Tampere University

TERHI RAVASKA

Essays on Optimal Lifetime Redistribution, Inequality and Well-Being

Tampere University Dissertations 75

Tampere University Dissertations 75

TERHI RAVASKA

Essays on Optimal Lifetime Redistribution, Inequality and Well-Being

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty Council of the Faculty of Management and Business
of Tampere University,
for public discussion in the Väinö Linna auditorium
of the Linna building, Kalevantie 5, Tampere,
on 14 June 2019, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Management and Business
Finland

Responsible supervisor and Custos	Professor Matti Tuomala Tampere University Finland	
Pre-examiner(s)	Professor Markus Jäntti Stockholm University Sweden	D.Soc.Sc Marja Riihelä VATT Finland
Opponent(s)	Professor Markus Jäntti Stockholm University Sweden	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-1119-3 (print)
ISBN 978-952-03-1120-9 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-1120-9

PunaMusta Oy – Yliopistopaino
Tampere 2019

ACKNOWLEDGEMENTS

My curious mind has always pushed me to ask questions and to study, but for a long time I doubted whether any of this could lead to anything of academic merit. This collection of essays shows what I have learnt during the past years. I feel tremendously thankful for all the people who have helped me to finish this project.

Professor Matti Tuomala supervised this thesis and I have learnt a great deal from him especially concerning the optimal taxation theory. I have also enjoyed the quite broad discussions about the world and our society with him. I have written two of the papers with Matti and I am grateful for all the knowledge shared. Sanna Tenhunen co-authored the first paper with us and Sanna gave me a crash course into the numerical simulations methods with Matlab. Thank you for your kind help. I thank Professor Jukka Pirttilä for occasionally monitoring what I was doing and putting me back on the right track and getting me more focused. I am also grateful that Jukka organized a short visit to the Uppsala University for me. I also thank my pre-examinators, Marja Riihelä from VATT and Professor Markus Jäntti from the Stockholm University for their helpful comments and suggestions.

I always felt welcome at the University of Tampere despite being very rarely present at the university. My colleagues and friends there deserve a big thank you for this. I thank Elina Tuominen and Jani-Petri Laamanen for their support and wisdom, my Tampere peer-group Matti Hovi, Allan Seuri, Heikki Palviainen, Sami Remes, Risto Rönkkö for the interesting conversations and my other colleagues Sinikka Hämäläinen, Jukka Ilomäki, Markku Konttinen, Professor Kaisa Kotakorpi, Hannu Laurila, Harri Nikula and Professor Jari Vainiomäki for contributing to the pleasant "Tampere spirit". Hannu also deserves the warmest acknowledgement for all the practical help he gave me during the past months.

I have been fortunate to work at the Labour Institute for Economic Research for the past three years. I am grateful to our former Director Seija Ilmakunnas for hiring me as a junior researcher and supporting me during the early steps of my career. I am also grateful to our current Director Elina Pylkkänen who has encouraged me to push ambitiously forward in my career. I want to express my deepest gratitude to our Research Director Tuomas Kosonen, who has relentlessly guided me towards good quality research.

I owe a debt of gratitude to Ilpo Suoniemi who has so many times corrected my texts and helped me with the substance of my research. Terhi Maczulskij has been an inspiring colleague and I feel grateful for all the advice and insight my namesake has shared with me, thank you Terhi for that. Hannu Karhunen has also been kind and shared his knowledge (and Stata code), thank you Hannu. Luncheons etc. with Esapekka, Henri, Lauro, Ohto, Petri, Sakari, Sami and Tuomo, among others, have brought much needed funny and instructive breaks during the hectic days. I have come to appreciate all the little anecdotes of seemingly unrelated things overheard during these moments. Cheers colleagues!

Besides finishing my thesis, for the past two years I have also been working and planning new projects with several collaborators. This has been the most motivating part. I thank Ohto Kanninen for taking me onboard in an interesting project during which I have learnt a great deal. The newest projects, one with Laura Peutere and another with Tuuli Paukkeri, have given me extra endurance to work long hours. Thank you Laura and Tuuli for committing to these projects and I am excited to finally allocate my time fully to these important questions with you.

I am grateful to the OP group research foundation for financing my first years in the FDPE program. I thank my FDPE friends who made the first years so pleasant. Specifically, I would like to thank Markku Siikanen who always made me feel a bit calmer during the first year of FDPE. I want to express my warmest gratitude to Krista Riukula who, beside being my occasional personal trainer, has given me valuable support in the moments of success but also during setbacks.

Two persons outside of economics departments deserve my deepest gratitude. Antti-Jussi Kieloaho happened to cross my path when I was at a crossroads. Thank you Antti-Jussi for sharing your own story so earnestly and giving me faith that I should at least try to finish this project. Carlos Diaz Ruiz has been my academic rockstar (or academic pirate) and I especially thank you for putting my research related worries in perspective. I also thank you Carlos and your beautiful Ksenia for your friendship.

I owe thanks to many friends for all the merry moments that kept me sane. Especially thank you Kristina Perälä for your friendship and reminding me of my roots, thank you Anna Oksanen for the fun moments and for bringing the music back into my life (significant improvement in my well-being) and thank you "Kansis-tytöt" Jenni Valkeapää, Marja Oksman, Saila Salonen and Helena Pulkki for our long-lasting friendship and for sharing your experiences in the professional life outside research. No matter what sort of glass-ceilings you might face, I am certain you will shatter them.

My father Veikko has taught me to work hard, my mother Rauni has planted the mind of an optimist in me, the Coloma family have proved, contrary to my own perception, that "a little" household chaos does not kill you and my sister Erja has always helped me to stay focused on the right things. With you in the background it has been easy to try to succeed. Thank you for your love and support. Finally, Joni, without you this would have been impossible. Thank you for everything.

On the first day of Spring, May 2019,
Terhi Ravaska

ABSTRACT

This thesis includes an introduction chapter and four essays on the field of public and welfare economics. The first two essays are theoretical and the last two essays are empirical. The theoretical part of this thesis studies the optimal taxation in a setting where the social planner has different kinds of redistributive preferences and individuals differ in more than one background characteristic. The empirical part of the thesis examines income distribution and the consequences of public policies. The first empirical essay studies top income distribution over time and the second empirical essay evaluates the possibilities for elderly workers to reduce work hours through a part-time retirement scheme and how this possibility affects sickness and drug purchases.

The objective of the public sector is to maximize social welfare by taking into account the efficiency of its policies and the equity perspectives. The main tool used is redistributing income through taxes and income transfers. The first essay of the thesis studies the optimal structure of income taxation in the Mirrleesian optimal taxation framework. The economy under study consists of individuals with different abilities to acquire income but they also have different preferences towards saving income for the future periods. The tax planner wants to redistribute income based on the ability to pay but the unobserved differences in saving preferences complicate the optimization problem. It is observed that in this kind of economy capital income taxation belongs to the optimal tax mix.

In the second essay optimal tax mix is studied in an economy where individuals differ with respect to their abilities and initial endowment or inheritance received. The tax planner cannot observe either of the factors. The goal is to redistribute income from those who have better ability and higher initial wealth towards the less skilled and less endowed. Several countries have abolished inheritance taxation and for this reason the starting point in this es-

say is that the tax planner can only use the non-linear labour income tax and, if necessary, capital income tax. It is shown that non-linear capital income tax belongs to the optimal tax mix. The essay also discusses the role of income shifting in this type of economy.

The effects of redistribution and other policies are seen as changes in the income distribution. The third essay in this thesis describes the Finnish income distribution and the evolution of incomes during the period of 1995-2012. The essay especially examines the top of the income distribution from a gender perspective. While the income shares of the high-income individuals grew rapidly at the end of the 1990s, this mostly benefited men. The share of women in the top incomes did not improve before the financial crisis apart from the top 1 percent, where the share improved throughout the period. Income mobility and income composition are also studied. Men's income ranks are more persistent than women's. Women had a bigger share of capital and business income but towards the end of the period the share of wages grew for women.

Beside the tax and transfer policy, the public sector can affect the well-being of citizens through different kinds of labour market and pension policies and programs. The fourth essay studies the part-time pension program and how it affected drug purchases and sickness. The part-time pension program reduced the hours worked but the combination of wages and pension reduced disposable income only by a little. The eligibility age for part-time pension programs was reduced in the year 1998, which enables the comparison of similar groups, where in one the reduction in work hours happened at a younger age than in the other. This research design makes it possible to evaluate the causal effect of the reform. In the study an instrumental variable method is also used to evaluate how moving to part-time pension affected drug purchases, sickness absences and labour market exits. The study finds that on average the age eligibility reform increased the purchases of drugs but for the sickest subset the reduction of work hours decreased sickness absences and drug utilization.

TIIVISTELMÄ

Tämä väitöskirja koostuu johdantoluvusta ja neljästä julkis- ja hyvinvoinnin taloustieteenalaan kuuluvasta tutkimuksesta. Kaksi ensimmäistä esseetä ovat teoreettisia ja kaksi viimeistä esseetä empiirisiä. Väitöskirjan teoreettisessa osiossa tarkastellaan optimaalista verorakennetta, kun julkisella vallalla on erilaisia tulonjakotavoitteita ja yksilöt eroavat useamman taustamuuttujan suhteen. Väitöskirjan empiirisessä osiossa tarkastellaan tulojakaumaa ja julkisen vallan politiikan seurauksia. Ensimmäinen empiirinen essee tarkastelee huipputulojen jakautumista ajan mittaan ja toinen essee arvioi ikääntyneiden työntekijöiden työn vähentämistä osa-aikaeläkeohjelman kautta ja sitä miten tämä vaikuttaa sairastavuuteen ja lääkeostoihin.

Julkisen sektorin tavoitteena on maksimoida yhteiskunnan hyvinvointia huomiomalla politiikan tehokkuus- ja oikeudenmukaisuusnäkökulmat. Yksi merkittävimmistä keinoista tämän tavoitteen saavuttamiseksi on tulojen uudelleenjako. Väitöskirjan ensimmmäinen essee tarkastelee vero- ja tulonsiirtojärjestelmän optimaalista rakennetta Mirrleesin (1971) optimituloverokehikossa. Esseessä tarkastellaan taloutta, jossa yksilöillä on erilaiset tulonansaintakyvyt sekä säästämispreferenssit. Julkinen valta haluaa uudelleenjakaa tuloa tulonansaintakyvyn perusteella, mutta havaitsemattomat säästämispreferenssit hankaloittavat optimointiongelmaa. Esseessä havaitaan, että tällaisessa taloudessa pääomatulon verotus kuuluu optimaalisten veroinstrumenttien valikoimaan.

Toisessa esseessä verotuksen rakennetta tarkastellaan Mirrleesin (1971) optimituloverokehikossa niin, että yksilöt eroavat tulonansaintakyvyn sekä alkuvarannon tai perinnön suhteen. Julkinen valta ei voi havaita kumpaakaan tekijää. Tavoitteena on uudelleenjakaa tuloa niiltä, joiden tulonansaintakyky on parempi ja joilla perintö on suurempi. Useat maat ovat poistaneet perintöverotuksen keinovalikoimasta ja tästä syystä esseen lähtökohta on, että julkinen valta voi hyödyntää vain epälineaarista työtulon veroa sekä mahdollisesti pääomatuloveroa. Esseessä näytetään, että pääomatulovero kuuluu optimaalisten veroinstrumenttien valikoimaan. Esseessä keskustellaan myös tulonmuunto-ongelmasta tällaisessa taloudessa.

Julkisen vallan uudelleenjakopolitiikan ja muun lainsäädännön vaikutukset näkyvät muutoksina tulojakaumassa. Väitöskirjan kolmas essee kuvailee Suomen tulojakaumaa ja siinä tapahtuneita muutoksia vuosien 1995-2012 välillä. Esseessä tarkastellaan tulojakauman huippua etenkin sukupuolten näkökulmasta. 1990-luvun lopun suurituloisten tulo-osuuksien kasvu kohdistui suurelta osin miehille. Naisten osuus tulohuipulla ei merkittävästi kasvanut ennen finanssikriisiä lukuunottamatta ylintä yhtä prosenttia, jossa naisten osuus kasvoi läpi periodin. Esseessä tarkastellaan myös suurituloisten tuloliikkuvuutta ja tulonmuodostusta. Miesten tuloasema huipputuloissa on pysyvämpi kuin naisilla. Naisten huipputulot koostuvat enemmän pääomatuloista mutta periodin loppua kohden palkkatulojen osuus kokonaistulosta kasvaa.

Vero- ja tulonsiirtopolitiikan rinnalla julkinen valta voi vaikuttaa kansalaisten hyvinvointiin erilaisten ohjelmien avulla. Esimerkiksi työ- ja eläkelainsäädännöllä voidaan tavoitella työntekijöiden parempaa hyvinvointia. Neljännessä esseessä tarkastellaan osa-aikaeläkejärjestelmän vaikutuksia lääkeostoihin ja sairaspoissaoloihin. Osa-aikaeläke vähensi työn määrää mutta käytettävissä olevat tulot pienenivät vain vähän. Vuonna 1998 osa-aikaeläkejärjestelmässä muutettin alinta ikärajaa, joka mahdollistaa muuten samankaltaisten ryhmien vertailun, joista toisessa työn määrää voitiin vähentää nuorempana. Tällainen tutkimusasetelma mahdollistaa ikärajareformin vaikutusten kausaaliarvionnin. Tutkimuksessa tarkastellaan myös instrumenttimuuttujamenetelmän avulla osa-aikaeläkkeelle siirtymisen vaikutuksia yksilöiden lääkeostoihin, sairaspoissaoloihin ja työmarkkinoilta poistumiseen. Tutkimuksessa havaitaan, että ikärajan alentaminen kasvatti lääkkeiden ostoja mutta sairaimmalle osajoukolle työmäärän vähentäminen vähensi sairaspoissaoloja ja lääkekäyttöä.

CONTENTS

ORIGINAL PUBLICATIONS

Chapter 2, "On the optimal lifetime redistribution and social objectives: a multidimensional approach", co-authored with Matti Tuomala and Sanna Tenhunen has been published in *International Tax and Public Finance*. The article is reprinted by permission from Springer. Published article is found at: Ravaska, T., Tenhunen, S. & Tuomala, M. Int Tax Public Finance (2018) 25: 631. https://doi.org/10.1007/s10797-017-9473-0 .

Chapter 4, "Top incomes and income dynamics from a gender perspective: Evidence from Finland 1995–2012" has been earlier published as a working paper: Ravaska, T. (2018). Top incomes and income dynamics from a gender perspective: Evidence from Finland 1995–2012. *Labour Institute for Economic Research working paper* 321.

1 INTRODUCTION

This thesis consists of four essays from the fields of public economics. Two of the essays study a traditional public economics question of optimal taxation. These essays study whether savings should be taxed when there are multidimensional differences between agents (instead of only one-dimensional differences in abilities or skills) and when we make different assumptions about the social welfare function. The essays ask a positive research question of whether savings ought to be taxed while keeping in mind the spectrum of normative criteria the government might set for welfare maximization.

The focus in public economics is social welfare maximization through government interventions. Income distribution is the most visible way of describing the consequences of government actions. The direct effect of tax-and-transfer system is transparent in the comparison of the factor and disposable income distributions. The shape and evolution of the gross income distribution over time indirectly reveal the effects of changing market structures and a broad set of policies implemented in the society. Analysing factors and characteristics behind these evolutions has been an important research field for the past decade. The third essay in this thesis describes the Finnish income distribution over time and focuses especially on the upper part of the income distribution from the gender perspective.

The fourth essay in this thesis studies the effects of policy on individuals' well-being and decision-making. The policy under study is the part-time retirement program which enabled eligible individuals to reduce work hours at the end of their career while the disposable income was barely affected. The essay studies the effects on health-related factors, such as drug utilization and sickness days, and early labour market exits.

These four essays add to the understanding in the field of public economics. The shared themes are equity and inequality, allocative efficiency and the well-being of individuals and society.

1.1 Optimal taxation and redistribution

The first two essays of this thesis consider optimal taxation. Optimal tax theory is interested in finding conditions for tax rates which maximize social welfare. The social welfare is represented with social welfare function and, by changing the form of this function, different normative concerns can be considered. Modern tax theory also acknowledges the important role of asymmetric information between the tax planner (principal) and the tax payer (agent). Information is a main element in modelling optimal taxes and distortionary taxes arise because of the information constraint. During the past two decades optimal tax problems have incorporated more heterogeneity between agents and this has lead to more realistic optimal tax models. This sort of multidimensionality requires numerical methods for solving the optimal tax rates and optimal tax structure. These building blocks of optimal taxation models are discussed deeper in the following subsections.

1.1.1 Optimal structure of income taxation

The basic theorem of welfare economics states that under a perfectly competitive economy and full sets of markets, the allocation equilibrium is Pareto-efficient. However, Pareto-efficiency does not entail that the distribution is in accord with the prevailing concepts of equity. One of the main activities of the government is thus redistribution[1]. Non-distortionary taxes and transfers are not available under redistributive preferences of the government and for this reason a trade-off between equity and efficiency needs to be considered, i.e. to what extent society is willing to suffer efficiency costs to obtain a more equal income distribution.

The optimal taxation problem is to maximize society's welfare by allocating bundles of consumption and leisure and using taxation as a tool to achieve

[1] Equity considerations are not the only reason for the case for redistribution. The common assumption, that the marginal utility of income is higher for the lower-ability type, suggest under aggregate welfare maximization that redistribution from high-ability individual to low-ability individual is desirable.

the optimal bundles. The modern theory of optimal taxation follows the seminal work by James Mirrlees where the trade-off between efficiency and equity and informational constraints are in a central role. In Mirrlees (1971) the government observes the income of individuals but the ability or efficient hours worked are not observed. On one hand the government wants to redistribute from the high-ability types towards the low-ability types for the sake of equity. However, if the government taxes too much, there is an incentive for the high-ability type to not reveal his true type. Instead, the high-ability type mimics the low-ability type for example by working fewer hours. This distorts the efficiency and reduces social welfare.

In these early contributions, optimal income taxation was considered in a static setting and the taxation problem focused on how to tax labour income. In the rigorous analysis of Mirrlees (1971) the factors for which the optimal tax rates are dependable were shown in general terms. These factors are the shape of the social welfare function, the self-selection constraints caused by the asymmetry of information, the distribution of ability and the production process. The analytical results from the early Mirrleesian framework taught that the marginal tax rate schedule is non-negative for most part of the distribution but zero for the highest ability type if skill distribution is bounded (Sadka, 1976; Seade, 1977) and zero for the lowest ability type if all individuals work non-zero hours (Seade, 1977). Subsequent work continued with the Mirrleesian framework with different assumptions about the skill distribution and noticed that the zero tax results hold only under restrictive assumptions.

Beside optimal labour income tax, an important question in taxation literature has been whether mixed taxation, a combination of commodity and income taxation, can achieve the same distributional objectives but with smaller efficiency costs. Some important early contributions in the field of commodity taxation are works by Ramsey (1927) showing that commodity taxes should be set to levels which cause a equiproportional shift in demands of each commodity, Corlett and Hague (1953) showing how optimal commodity taxes are higher for products which are complementary to leisure and Diamond and Mirrlees (1971a,b) who pondered the production efficiency under linear commodity taxation concluding that it is optimal to operate at the production-possibilities frontier also under distortionary taxation.

Atkinson and Stiglitz (1976) combined the commodity taxation and income taxation in their model, which in subsequent literature is known as the Atkinson-Stiglitz result. Atkinson and Stiglitz (1976) show that under optimal[2] non-linear labour income taxation there is no need for indirect taxation when preferences are separable between goods and leisure. Savings can be thought of as a commodity since savings are consumed in the future period. The Atkinson-Stiglitz result implies that savings should not be taxed for reasons of redistribution. Later research on optimal capital income taxation has presented various interesting cases where the Atkinson-Stiglitz result does not hold[3].

In the case of indirect taxation, the Atkinson-Stiglitz result does not hold[4] if production technology is non-linear (Naito, 1999), when there are differences in the unobserved endowments (Cremer et al., 2001), differences in preferences (Saez, 2002; Marchand et al., 2003; Blomquist and Christiansen, 2008) or when there is wage uncertainty (Cremer and Gahvari, 1995a,b). With the direct application to capital income taxation the Atkinson-Stiglitz result is shown not to hold with multidimensional individual heterogeneity which is unobservable to government for example in the case of habit formation (Tuomala and Tenhunen, 2013) and in the case of preference heterogeneity (Saez, 2002; Tenhunen and Tuomala, 2010; Diamond and Spinnewijn, 2011; Ravaska et al., 2018). Also if there is a possibility to shift income between capital and labour income tax base, capital income tax can be welfare-improving (Christiansen and Tuomala, 2008). The role of capital income taxation in these settings is to mitigate the information asymmetry between the agent and principal. Stiglitz (2018) also discussed the Atkinson-Stiglitz result in a dynamic model and finds that there is a case for non-zero capital income tax.

[2]Kaplow (2006) showed that optimality of the labour income tax is not required for the Atkinson-Stiglitz result to hold.

[3]The zero capital income tax result for the steady state is also derived by Chamley (1986) and Judd (1985) under assumptions of representative individual who lives infinitely or when there is an infinite dynasty. If in the Chamley-Judd setting households face tight borrowing constraint or are subject to uninsurable idiosyncratic income risk, the zero-tax result is invalidated (Aiyagari, 1995; Conesa et al., 2009).

[4]The main assumption behind the Atkinson-Stiglitz result is that preferences over goods demand and labour supply are weakly separable. However, there is empirical evidence that this might not be the case (Browning and Meghir, 1991; Crawford et al., 2010; Pirttilä and Suoniemi, 2014).

The optimal tax theory suggests that the marginal income tax rates should be higher for products and factors which have small behavioural elasticities. Also, conditions in the tax problems clearly show that marginal tax rates are higher if the government has stronger preferences for redistribution and for the parts of the income distribution where there are fewer individuals. The current knowledge from the theory of mixed optimal taxation points to the direction that labour income should be taxed non-linearly and there should also exist a non-zero tax on capital. The optimal tax theory does not rigorously answer what the relation between marginal labour tax rates and marginal capital income tax rates is. There are reasons to believe that the tax rates should not be the same but to minimize income shifting they should be somehow related (Banks and Diamond, 2010).

1.1.2 Social welfare function

In economics, social welfare is characterized with a real-valued welfare function which depends on variables that affect collective well-being. The normative part of the theory of optimal taxation can be reduced to the choice of the social welfare function. The social welfare function expresses the society's values about equity and fairness. The common way of thinking about social welfare is with the Bergson-Samuelson social welfare function where social welfare is aggregated in a certain manner from the individual utilities (Tuomala, 2016). The objective of the society is to maximize this aggregated welfare. This approach to social welfare is called welfarism. Equity considerations are incorporated in the functional form of the social welfare function. The commonly assumed property of the social welfare function is that it is increasing in everyone's utility and so it entails the Pareto principle.

Utilitarian social welfare function includes all individuals in the economy while under Rawlsian or the maximin principle the objective function includes only the utility of the worst-off individual (or group). Different kinds of redistributive preferences can be represented with the choice of social welfare weights schedule[5]. Beside maximizing the chosen objective function, the

[5]In the last few years there has been an increase in the empirical applications which aim to reveal the underlying social preferences with the help of the so called inverse optimal-tax approach. Here the assumption is that the observed tax rates are optimal and then the optimal

policy maker needs to consider the constraints in the optimization problem. An important set of constraints are the self-selection constraints,[6] which incentivize individuals to reveal their true types instead of mimicking another type. While the social objective defined as the maximin case does not include the utility of the higher ability types directly in the objective function, naturally these ability types need to be considered in the constraints.

The social welfare can also express other value judgements beside welfarism. In these cases the social welfare is not derived from individual preferences. For example, in the context of developing countries, poverty-minimization is a common non-welfarist objective of the government where the social welfare weight is zero for the income groups whose incomes are above the level of poverty. Other types of non-welfarist objectives are based on inequality indices or capabilities.

There is much debate whether social objectives should equalize the utility differences in the welfarist sense or aim towards reducing relative poverty. Even when redistribution is dominantly preferred, the debate often gets stuck on the welfare weights for different groups. While some argue for the utilitarian case as fair and sufficient option for the form of objective function, others debate for the maximin case, that is focusing on the groups who are the most vulnerable. While this normative debate cannot be solved with the tools of economic theory there is a more attractive and less debated policy goal available, namely the equality of opportunity.

Equality of opportunity, the idea that individuals should have equal opportunities to support their own living, is popular among philosophers, politicians and the general public. The equality of opportunity approach has roots in the claim by Dworkin (1981a,b) that egalitarian redistribution should not be implemented based on certain kind of preferences. Roemer (1998) formulated the same idea by stating that factors influencing the outcomes should be partitioned in terms of effort and circumstances and to recognize that some income differentials are due to factors of choice and others on factors of circumstances. The individual choices, when assumed they are independent

tax formulas can be used in the inverted form to calculate the respective social welfare weights schedule. This method has also been applied to evaluate the social welfare weights of political parties towards different groups. See Jacobs et al. (2017) and the references therein.

[6] Also known as the incentive-compatibility constraint.

of circumstances[7], do not require governmental intervention in the form of redistribution. However, the income differences based on the factors of circumstances that the individual cannot affect himself should be compensated. These two points are known as the principle of responsibility and the principle of compensation (Roemer, 1998).

The principles of compensation and responsibility can be characterized with an example of individuals who differ in their innate ability but also have different preferences for work. This is realized as differences in income also at the same ability level. While the ability differences are not under an individual's control, the government should redistribute from a high-ability type towards the low-ability type. However, it is often assumed that effort or work preferences are something that an individual can make active decisions on. So, there is no equity argument for redistributing from the hard-working towards the individual who puts less effort in work. Neither is there reason to penalize the individuals who have lower effort as preferences are valued on their own without moral judgements.

The equality of opportunity goal makes the optimal tax problems effectively multidimensional. In recent years there have been several attempts to empirically characterize inequality in terms of equality of opportunity (Roemer and Trannoy, 2016; Ramos and Van de gaer, 2016). These applications face difficulties because there is no clear distinction of which factors belong to the group of responsibility of an individual and which are circumstances. Theoretically one way to incorporate equality of opportunity is to put different weights for each group, even when there is no natural way to weight preferences (Boadway et al., 2002).

A more agnostic way to incorporate equality of opportunity in the social welfare function is by making a compromise between the two guiding principles of responsibility and compensation. Roemer (1998) and Van de gaer (1993) suggest that equality of opportunity social welfare function combines the elements of maximin and utilitarian cases. The maximin element steps into the picture in the form of high-inequality aversion applied along the dimension of circumstances and the utilitarian element shows up as the zero-

[7]For example Kanbur and Wagstaff (2014) point out that a clean separation between circumstances and effort is not always possible even conceptually. This observation implies that the principles of responsibility and compensation are not adequate guidelines.

inequality aversion in the dimension of responsibility. This approach was incorporated in the analysis by Ravaska et al. (2018). Fleurbaey and Maniquet (2018) discuss extensively how other fairness principles can be incorporated in the social welfare function.

1.1.3 Multidimensional optimal tax problems

The traditional Mirrleesian optimal tax models have assumed that individuals in a given society have similar preferences between consumption and leisure, but the innate ability, which is perfectly reflected by the wage rate, differs. These preferences can be represented with the same utility function and this has made the analytical solutions interpretable.

It is highly plausible that individuals differ in their preferences and/or also in other relevant (for taxation) background characteristics which should be taken into account in an optimal taxation problem. Individuals have for example different work preferences, different timing preferences for consumption and are endowed with different resources in the beginning of their life. This type of multidimensional heterogeneity adds reality to the optimal tax and transfer problems but at the same time complicates the analysis notably. Increases in computational power have made it possible to solve these more complex models numerically.

The technical difficulties caused by multiple dimensions of private information are discussed in Armstrong (1996) and Rochet and Choné (1998). The context in these papers is multiproduct nonlinear pricing schemes in microeconomic mechanism design problems. A similar kind of solution algorithm has also turned out useful in the optimal taxation problem but with the additional resource constraint that the government poses to the problem. Armstrong and Rochet (1999) develop a simplified version of the multidimensional screening problem where they consider the simplest non-trivial problem, which is with discrete type space and where each type parameter comes from a binary distribution. This type of tractable model can provide a complete solution and has also been used in the optimal taxation literature for example by Cremer et al. (2001), Tenhunen and Tuomala (2010), Diamond and Spinnewijn (2011), Tuomala and Tenhunen (2013) and Ravaska et al. (2018).

The heterogeneous preferences and other type of multidimensional variability between individuals complicate the optimal tax analysis from at least two points of view. Firstly, the common logic of incentives need not to work anymore. For example, if individuals at the same ability level value work differently, a policy maker cannot offer similar incentives to get these types to work the same amount. The derivation of multidimensional taxation problem becomes more difficult as there are more self-selection constraints to account for. The analytical results seldom reveal the binding constraints and for this reason the analytical derivation of the problem might be fruitless. However, numerical simulation can reveal much more about the optimal tax-and-transfer system.

The second complication is that, with more complex heterogeneity, defining a social objective is more difficult. In the case of homogeneous preferences, the objective function makes interpersonal comparisons based on utilities differing by only the work effort (which depends on hours worked and productivity). Interpersonal comparisons of heterogeneous preferences is conceptually difficult in the welfarist tradition since the interpretation of these utilities is not impartial anymore.

Heterogenous working or consumption preferences have been introduced in the optimal taxation model for example by Boadway et al. (2002), Cremer et al. (2009), Tenhunen and Tuomala (2010), Diamond and Spinnewijn (2011), Tuomala and Tenhunen (2013), Jacquet et al. (2013), Golosov et al. (2013), Lockwood and Weinzierl (2015) and Ravaska et al. (2018). These papers have solved the technical complexities differently. For example Boadway et al. (2002) assumes a utilitarian social welfare function so that the government chooses different weights for individuals with different preferences. They also simplify the analysis by assuming the direction of binding self-selection constraints. Cremer et al. (2009), Tenhunen and Tuomala (2010) and Tuomala and Tenhunen (2013) on the other hand rely on numerical simulations and remain agnostic about the appropriate cardinalization of the different preferences. Some papers avoid the technical difficulties by assuming multidimensionalities can be represented with one-dimensional aggregation or that there is perfect correlation between dimensions which also reduces the problem to unidimensional (Lockwood and Weinzierl, 2015; Golosov et al., 2013; Choné and Laroque, 2010). In the case of capital income taxation one simplification has also been to consider linear capital income tax or subsidy[8] (Diamond and Spinnewijn, 2011).

1.1.4 Numerical methods

In the one-dimensional optimal income taxation model, analytical results provide qualitative characterization of the optimal shape of the marginal tax schedule while numerical simulations can give indications also towards the levels of tax rates. However, the difficulty to derive exact levels of optimal tax rates is that much of the model parameters are unknown and hard to estimate empirically. For example, optimal tax formulas include a parameter for labour supply elasticity with respect to tax rate. These elasticities have been estimated from specific quasi-experiments but in the end the external validity toward general taxation purposes is debatable.

When tax problems get more complex, like in the case of multi-dimensional heterogeneity between agents, the analytical results cannot always even reveal the sign of the optimal tax rates. In the more complex models the only way to get intuition on the optimal structure of the system is to solve them numerically with reasonable assumptions. Naturally the multidimensional problems also face the difficulty that many of the parameters of the model are unknown empirically.

For numerically deriving the optimal tax rates, one needs to specify four key elements: an individual's preferences, the shape of the ability distribution, social objectives and the revenue requirement. Social objectives and the choice of welfare function was discussed in the section 1.1.2. The individual preferences, presented in the form of the utility function, is a central question because different utility functions imply different types of behavioural responses. In the standard models the labour supply decision is based on the balance between after-tax income, leisure and consumption. Taxation decreases effective wage and increases the amount of work through the income effect. On the other hand, substitution effect decreases work amount as leisure

[8] Diamond and Spinnewijn (2011) also assume a four-type setting where only high-ability types choose between jobs instead of hours resulting to reduction in the number of possible binding self-selection constraints.

is cheaper than without taxation. The total effect depends on the utility function. In the optimal taxation literature various forms of utility functions have been tested.

The concept of distribution in optimal tax models refers to the ability distribution as discussed in the previous sections. However, this is unobservable. Assuming competitive labour markets where the wages are determined according to the marginal productivities, the wage distribution would work as an ability distribution. However, we do not observe the counter-factual distribution which would occur without the intervening role of taxation. The simplest way to get around this is to consider discrete types and a uniform distribution (for example Stern (1982); Stiglitz (1982)). Mirrlees (1971) considered log-normal ability distribution in the numerical calculations, and the subsequent analysis have for most part followed this approach. The log-normal distribution fits well in most parts of the empirical earnings distribution but the fit at the bottom and at the top of the earnings distribution is not so good. The empirical top of the earnings distribution has heavier tail than the log-normal distribution would indicate and at the bottom we see bunching.

Instead of fixing the distribution beforehand, there are also methods developed for continuous cases to derive the underlying ability or skills distribution from the empirical income distribution. These sorts of calibration exercises have been done by at least Saez (2001) and Kanbur and Tuomala (1994). Saez (2001) backs out the ability distribution from the empirical earnings distribution by assuming that the elasticity of labour supply is constant and approximates the actual tax schedule with linear tax. An alternative approach was presented in Kanbur and Tuomala (1994), where non-linearities of tax schedule is accepted together with utility functions which do not imply constant labour elasticities. The skill distributions are chosen so that the model produces the empirical distribution.

An example of an efficient numerical computing environment for finding the solution for optimization problems is Matlab. The ready made algorithms, such as *fmincon*, solve multi-variable functions with constraints that may be linear or non-linear and also inequality constraints can be included. The program determines the binding inequality constraints so no a-priori assumptions need to be made. This is a convenient characteristic for the multidimensional optimal taxation applications since there is no theoretical guiding for which of the constraints should be slack and which are binding. However, researchers often also face technical difficulties in the numerical simulations, for example that for some parameter values the problem is not solvable. For this reason a set of sensitivity analysis needs to be conducted.

1.2 Top incomes

The third essay of this thesis focuses on the top incomes. While previously the literature on economic inequality concentrated on documenting and explaining the evolution of poverty, in the last decade inequality research has focused more on top incomes. Top incomes have increased rapidly in most developed countries which makes this field one of the most dynamic research areas in inequality literature. Studying top incomes in detail can also help interpreting the evolution in the overall inequality. Top incomes affect the outcomes in the other parts of the income distribution either through markets or through political decision-making.

The shape of the income distribution has changed and the gap between the very rich and the middle class is getting larger. The recent years, the studies on top incomes have tried to find explanations for the evolution of top incomes and broaden the literature also to include new perspectives, like the issue of gender. In the following subsections I briefly summarize the current knowledge on top income shares, the research on the role of income mobility and introduce the literature of gender and top incomes.

1.2.1 Top income shares: international perspective

Kuznets and Jenks (1953) constructed first top income shares series. They used the US income tax tabulations and computed the top decile income shares series for the years 1913-1948. Based partly on these earlier findings, the famous Kuznets curve -hypothesis was articulated in Kuznets (1955). This hypothesis states that income inequality follows an inverted U-shape along the development: economic growth would first raise the inequality and through reallocation of workers to high-productivity sector, the inequality would eventually decline. When top income shares are used as an inequality measure, this hypothesis implies that in the beginning of economic development the income concentrates at the top as they are driving the development with their skills and productivity but in the long-term, as more people become educated and more productive, the income shares at the top would decrease.

The theme of income distribution was mostly neglected by the economic profession until the end of 1990s when it was brought "back in from the cold" as suggested by Atkinson (1997). The international interest in top income literature was revived by Piketty and Saez (2003), who studied the evolution of top incomes in the US since 1913. They showed that the share of total annual income received by the top 1 percent had more than doubled from the 1970s to 20 percent in 2011 indicating a reversal of the Kuznets hypothesis. This rise has significant effects on the overall income inequality as well (Atkinson et al., 2011, pp. 10). After Piketty and Saez (2003) top income shares series have been constructed systematically all over the world. Many of these studies are collected in the volumes edited by Atkinson and Piketty (2007, 2010). Their collection of data and studies aimed to provide a comparable, long-run and high-quality data source on income distribution. The time-series constructed for these volumes and for the subsequent database[9] utilize similar methodology cross-country making the data applicable for studying the explanations behind the long-run inequality.

Development in the top income shares is in many aspects similar in the Anglo-Saxon countries until the 1980s (Piketty, 2007; Alvaredo et al., 2013). There was first a secular decline in income inequality during the period of 1914-45. In the 1950s-60s inequality kept declining but at a slower pace in a number of countries while in the post-1970s the divergent cross-country pattern emerged. The US saw a rapid increase in the top income shares, the UK and Germany a milder increase and in continental Europe the shares remained fairly stable. Atkinson and Piketty (2010) included also Asian and Nordic countries into the analysis. While in the Nordic countries the economic inequality has been and still is very low, in the 1990s inequality growth was faster than in other European countries. The explanation provided is that while the wage income is relatively equally distributed, the capital income is

[9]World inequality database, https://wid.world

distributed more unequally (Jäntti et al., 2010; Roine and Waldenström, 2010; Aaberge and Atkinson, 2010).

To gain a better understanding of the development of top incomes there has also been a renewed interest in the decomposition of income inequality into labour income and capital income components. Especially at the top, interest income, rents and dividends form a significant proportion of the total personal income. The economic mechanisms behind these two components can be different, though intersected, and thus it is important to observe them separately in the attempt to find explanations for rising top incomes. Labour incomes are affected by the demand and supply of different kinds of skills as well as labour market institutions and the bargaining power of the workers, while the distribution of capital is more affected by credit constraints, capital accumulation and wealth taxation. The declined trend of income inequality in the period of 1914-45 was mostly driven by the fall of capital income. On the other hand, the enormous increase in the US top income shares in the last three decades has been more driven by increasing top executive wages while in Europe the capital income has been a more important source of growing top incomes. (Piketty, 2007, p.11).

Recently, with data matching business profits with the business owners, Fairfield and Jorratt De Luis (2016), Wolfson et al. (2016) and Alstadsæter et al. (2016) have shown that it is important to account for profits inside a firm when studying top income inequality. These accrued gains or business profits are often neglected due to data restrictions. As top income shares are often calculated from the individual-specific tax registers, which exist for taxation purposes and not originally for research purposes, the picture of top end inequality is underestimated. Personal tax registers report realized capital gains which are very volatile for several reasons, one of which is changes in reporting behaviour due to tax reforms. When the retained earnings of firms can be allocated to owners we gain a better understanding of the levels of top income shares and changes in the personal top incomes[10]. However, the ownership structures of different kinds of firms are complicated and there is a need for improved data in order to fully allocate the accrued profits to owners.

[10]For Norway allocating the corporate profits to shareholders more than doubles the top 0.01% income share (Alstadsæter et al., 2016).

Various explanations behind rising top incomes have been provided. For example technological development have created polarized labour markets where there is a low supply but high demand for high-productivity workers, which increases the returns at the very top of the income distribution (Autor, 2014). Alvaredo et al. (2013) however doubt that this cannot be the sole explanation because high-income countries have similar technological and productivity development but the patterns of top incomes vary. They propose that institutional and policy differences play a role. They show that there is a strong correlation between the reductions of top marginal income tax rates and the increases in the top 1 percent share of total income[11]. Other explanations for the rise in top incomes are the improved bargaining power and rent-shifting possibilities of the top executive (Kleven et al., 2014; Bivens and Mishel, 2013) for example due to globalization, increases in performance pay (Lemieux et al., 2009), deregulation in the financial sector (Philippon and Reshef, 2012) and increases in the innovation intensity (Aghion et al., 2019). Also, in line with "superstars" theory (Rosen, 1981)[12], increasing inequality in a few sectors can spill over to other sectors which then affects the overall inequality at the top (Clemens et al., 2016).

Especially in Europe, the private wealth to national income ratios have grown since the 1980s (Piketty and Zucman, 2014). The bequests, inheritances and gifts *inter vivos*, have returned as a source of inequality. The inheritance flows are scarcely available in the datasets but indirectly some part of capital income represents bequests. The most recent contribution to top income literature has been to study the association between capital and earnings distributions with copulas[13] (Aaberge et al., 2018). The joint distribution is an important future research avenue for two reasons. Firstly, it is important

[11] Marginal tax rate reductions can increase the top incomes by incentivizing the top managers to bargain for higher wages and away from compensation in the form of company perks (Piketty et al., 2014). Tax avoidance and evasion is another explanation for the negative correlation between the top marginal tax rates and top income shares. The change in top incomes can so partly stem from the tax avoidance behaviour instead of true changes in the concentration of income (Slemrod, 1996).

[12] Superstar theory explains how small differences in skill result in large differences in income.

[13] The rising top incomes can stem from rising earnings, rising capital or the increasing association of the two marginal distribution. Copulas are functions that define the bivariate distributions, here total income distribution, with the ranks of the marginal distributions of capital and earnings income separately.

to understand whether the working-rich will strengthen their position in the top incomes through capital and wealth accumulation. This amplifies the inequality process further. A second important research question is to find out whether bequests and high capital shares when young are associated with top labour earnings in prime-age, and what are the role of family wealth, family connections or intergenerational skills transfer in this process.

1.2.2 Income mobility at the top and lifetime income inequality

For a long time the information we had about inequality was based on annual cross-section data. This data restriction made it impossible to account for income mobility which is crucial in order to understand the patterns of more permanent inequality. The expansion in the availability of longitudinal datasets has increased the number of studies taking into account income mobility. Nowadays in many cases we can follow the same individuals for decades, and in some cases from the cradle to the grave.

Averaging individual income over several periods reduces some volatility of incomes and so reveals a more permanent inequality within a society. This measure is indicative of long-term inequality but is limited to describing it at the aggregate level. Income mobility measures reveal whether individuals move downwards or upwards in the income ranks over time.

Mobility can be studied within a person's lifetime (*intragenerational mobility*) or between generations (*intergenerational mobility*). Shorrocks (1978, pp. 378) defines intragenerational income mobility as "the extent to which the income distribution is equalised as the accounting period is extended". The chosen time period matters in measuring income mobility. Short-term intragenerational mobility is captured when this year's income position is compared to the next year's income position while a longer term mobility measure can include the whole lifetime or some other timespan. Intergenerational mobility looks at how a parent's and their children's income positions are associated. The multiple ways of measuring income mobility is summarized in Jäntti and Jenkins (2015).

Income mobility is preferred[14] as it signifies a more dynamic economy with

[14]Mobility is not socially desirable if it only represents transitory shocks in income, that is

social mobility and meritocracy compared to the case of no mobility[15]. Annual income inequality is less of a concern if there are mechanisms which make it possible for a low-income family member to move up in the income ladder. That is, income mobility is preferred for its instrumental reasons while on its own it is not important (Jäntti and Jenkins, 2015).

One mobility measure is to see how persistently individuals stay in a certain income group over time. Individuals at the top of the income distribution have more transitory income ranks compared to the middle of the distribution. For example, reaching the top 0.1% requires extremely high incomes which are transitory in nature stemming from the selling of a firm or winning the lottery. Only a few individuals can maintain extremely high earning capacities year after year. Rather than pointing out the level of persistence, focus should be put on the evolution of persistence as this reveals the trend in permanent inequality and mobility. For example Kopczuk et al. (2010) report that the probability to stay in the top 1 percent in U.S has changed little over the past decades.

The interpretation of top income mobility with general mobility measures is somewhat more limited because by definition from the top groups one can move only towards the lower income ranks. Also the group sizes vary if a researcher focuses solely on the top and looks at mobility within the top decile, top 1% and so forth. In these cases it is better to divide the top into equal group sizes or use other more general measures, for example the income mobility curve (Aaberge et al., 2013). Aaberge et al. (2013) apply the income mobility curve for the top incomes and take into account the extent of income changes together with changing income ranks. The former point is especially important in cross-country comparisons because in a country with more equal income distribution, a small income increase can change the income rank substantially and thus show high mobility.

The tax reforms that induce changes in income reporting affect mobility measures. Just like with calculating the top income shares, the preferred in-

---

income uncertainty, which has a welfare decreasing effect (Shorrocks, 1978).

[15]Immobile society would be a consequence of an economy where incomes perfectly represent ability year after year and transmission of skills between generations were perfect. However, it is more plausible that for example institutions and public policies matter because there are large differences in mobility between similar countries. For a comparison of the US and Canada see Corak (2013).

come concept include all income sources in order to get a full picture of the persistence of the top income receivers. A study by Alstadsæter et al. (2016) note that linking business profits or retained earnings to the owners stabilized the movement out of the top income groups. This indicates that part of the mobility observed in the personal tax base can be caused by the responses to changes in the legislative environment.

The intergenerational mobility of top incomes is a less studied topic because the datasets linking generations and covering the top are limited. However, this is possible for few countries. For Canada and the US, Corak and Heisz (1999) and Corak (2013) illustrate that there are non-linearities in the intergenerational mobility. Sons whose fathers are in the top decile are more likely to be top income receivers in adulthood, more so for the US, and this correlation is stronger at the very top. The strong link from a parent's earnings rank to a son's earnings rank can be due to skill transmission through genetic factors, or indicating better education, employment opportunities and networks for the children of the wealthy parents. Corak and Piraino (2011) document that for Canada 7 out of 10 sons shared the same employer as their father who was in the top 1% of the earnings distribution. This raises a question whether the sons have an advantage because of acquired firm-specific skills from their father or whether nepotism is taking place.

Non-linearities of income mobility are also shown for Nordic countries in Bratsberg et al. (2007) and Suoniemi (2017). With Swedish data Björklund et al. (2012) can focus especially on the intergenerational mobility of the top 0,01%. While the Swedish intergenerational mobility is high in general, the mobility at the very top is weaker. The correlation between the father's and son's total income rank is especially strong for the top income distribution, nevertheless the mobility gets weaker at the top of the labour earnings distribution as well. The authors also study the mechanism behind the income transmission and it seems that wealth is a likely explanation instead of skills measured as IQ, education or non-cognitive skills.

1.2.3 Gender and the top income distribution

Wage inequality is a traditional gender economics question. We know that on average the raw wage gap has reduced over time and controlling for differences in schooling, work experience, industry, occupation, union status and hours worked reduces it further. However, there is still a large unexplained part in the total wage gap, which has remained fairly stable. For the top of the earnings distribution, it has been observed that pay gap has declined much more slowly than for the average worker and the unexplained part is largest at the top of the wage distribution. (Blau and Kahn, 2017).

The latest addition in the studies of high-skilled women and men has been to expand the view from earnings to total incomes. Total incomes include, alongside wage income, self-employment income, capital income and transfer income. In recent years there have been attempts to characterize the top income distribution also from a gender perspective. It has been observed that under-representation of women at the top is quite a common phenomenon on across developed countries and there are clear gender differences in the income composition and income mobility. (Roine et al., 2017; Atkinson et al., 2018; Ravaska, 2018)

As the largest fraction of individual total income comes from wages, the under-representation of women in the total income distribution can be partly derived back to the wage gaps. Despite the advancements in labour market progress over the last decades, women are under-represented in high-earnings and high-status occupations. A recent study suggests that much of the overall gender pay gap can be explained by the missing women at the very top of the earnings distribution (Fortin et al., 2017). Despite the persistent under-representation in both the top income and top earnings distribution, the women's share has improved over time (Guvenen et al., 2014; Roine et al., 2017; Ravaska, 2018; Atkinson et al., 2018).

Many different explanations exist for the gender wage gap and for why women are not well represented in the upper part of the earned income ladder. The literature related to this is well summarized in Blau and Kahn (2017). Concerning the top incomes, the under-representation of women might occur because the women are newcomers in the high productivity jobs. It takes time to move to the company's boards and manager positions and accumulate enough savings from earnings to accrue high capital incomes. This explanation is called a pipeline argument in the gender wage gap literature. Another explanation is that there is an invisible barrier, the so called glass ceiling, which could occur because discrimination or other more subtle barriers which either make women less productive or less eager to get to the top positions. For example, combining family and household work with market work might push women to pursue less ambitious careers[16]. According to the evidence from the US this is true for an average high-skill woman since availability of substitutes for household production increased the number of women entering occupations with high returns for long working hours. At the top decile, however, the availability of substitutes for household work did not affect women's career progress. (Cortés and Pan, 2019). This indicates that there are other factors beside working hours affecting women's move to the top. One such factor could be differences in preferences concerning career paths and success but the empirical evidence is still lacking.

Public policies have aimed to improve the career possibilities of women through three instruments. Firstly there are "the equal pay for equal work" initiatives to abolish gender based discrimination as well as codes to promote gender equality. The second important instrument concerns family policies in the form of child-care services. The third instrument, more directly applicable to the top of the skill distribution, is gender quotas or voluntary codes promoting higher female representation in the boards of companies. While the two first instruments have improved the overall representation of women in the labour markets, the effect for the highest skill-level is unclear. The effect of gender quotas or codes is still ambiguous, however the evidence from Norway is discouraging: the binding gender quotas have not affected the overall gender pay gap or the representation of females in other parts of the income distribution than at the very top (Bertrand et al., 2019).

Beside the earnings differences between genders, the differences in invest-

[16] A related observation is that there is a high and increasing (over time) return to working long hours or particular hours of the day. Due to household work and childcare, women are more likely to be in a disadvantaged position in occupations where long working hours are required for high returns. Industries which have incorporated more workplace flexibility have also witnessed a shrinking gender wage gap. (Goldin, 2014).

ment decisions and opportunities show up as varying capital income. Atkinson et al. (2018) observe that in Norway the association between being at the top of the capital income distribution and top of the earnings income distribution is much stronger for men. This observation can stem from the mechanism that women at the top of the income distribution tend to inherit the wealth which generate top incomes (Edlund and Kopczuk, 2009). However, the gender differences in investment and capital income are still under-studied and not much can be said. This would be a fruitful future research avenue.

1.3 Effects of work hours on well-being

A principle in economic modelling is that an individual's well-being increases with consumption of goods and leisure and decreases with hours spent at work. In health economics the workhorse model by Grossman (1972) on health demand starts by assuming a trade-off between investing time in one's health and allocating it elsewhere. These two modelling frameworks suggest that work hours and ill-health are positively associated. However, the empirical applications on this relationship are scarce. The fourth essay in this thesis evaluates this relationship in the context of the elderly workforce.

Analysing the health or well-being effects of work are difficult because of the asymmetry of information. Individuals have heterogeneous preferences towards work and choose occupation and work hours accordingly. Hours spent at work affect health differently in different occupations. Also there is reverse causality: health affects the work hours decision. To estimate the causal relationship, either an experiment or a quasi-experiment is required. In the next subsections I will discuss the literature on the theoretical modelling of health and estimating the causal effects in this context.

1.3.1 Health stock model

The human capital model assumes that the investments in knowledge raise productivity and this leads to increases in the monetary returns in the labour markets. For this reason, an individual has an incentive to allocate time and resources into education, either in the form of formal schooling or in on-the-job training. If the expected returns are higher than the direct costs and the opportunity cost of time spent studying, the individual invests in acquiring education. The optimal quantity of investment varies in different phases of life and between individuals of the same age but with varying abilities. A similar kind of modelling can be extended to health.

We can think of "good health" as a commodity which is demanded. Each individual has an initial stock of health when they are born. The health stock produces healthy time which increases well-being. However, this health stock is depreciating with age but an individual can make investments which increase the stock. Investments are not costless and their (shadow) price is assumed to be increasing with age. This leads to different amounts of health (and medical care thereof) demanded at different points of the life-cycle. The shadow price for health is affected by many factors, like the price of medical care and the cost of exercise. (Grossman, 1972).

In this modelling framework initiated by Grossman (1972) health is demanded because it is on one hand a direct source of utility but also an investment commodity which makes more healthy time available for market and non-market activities. Unlike in the human capital investments, the increases in the stock of health does not only affect the productivity and the wage rates but health also affects the time constraint one can spend in market or household production. This is an important difference between the health stock model and the human capital formation. As an example, an individual working as an economist is usually considered a high-productive worker who has invested a lot in the human capital formation. However, the initial health stock and the depreciation rate of this stock affect how well those returns can be collected in the form of wages. The economist's number of projects conducted are dependent on this health constraint.

The work itself can also affect the stock of health and the depreciation rate. While this latter point is more discussed for example in the field of sociology, economists have just recently started to explore the effects of work on health (see chapter 5 and references therein). In the Grossman model the total amount of time in any period is divided between time at work (market production), time at leisure, time producing health and time lost from market and non-market activities due to sickness. If the time producing health increases, for example because work hours are reduced, keeping other inputs like income fixed, the model predicts improving health. However, one can also argue that reduced working hours can increase leisure activities which deteriorate health. To my knowledge, this type of mechanism has not been incorporated in the Grossman model. Finally, this is an empirical question.

1.3.2 Causal inference between work hours and health

It is particularly challenging to identify causal effects between work characteristics and health. Work characteristics are endogenous to the underlying health stock. In the example of work hours' effect on health, there is reverse causality when health affects the decision of working hours. Experiments or quasi-experiments are required in order to reveal the causal effects between work and health.

The methods for policy evaluation are reviewed in Abadie and Cattaneo (2018). In the ex-post programme evaluation the goal is to find exogenous variation. If a randomized experiment is run, exogenous variation comes directly from the setting and treatment effect comes from the comparison of the treatment and control groups. Unfortunately, running randomized experiments is seldom available and sources of exogenous variation need to be looked at from the social environment.

One popular method in programme evaluation is using a difference-in-differences approach. To use this method we need to find a treatment group, whose work hours changed exogenously, and a control group, who do not experience any shock in their working hours. For example such a setting took place in France in the 1990s when statutory weekly work hours were reduced by employer characteristics (Bietenbeck and Berniell, 2017).

An important question in the difference-in-differences setting is to what extent the control group is credible[17]. Naturally, we want to compare individuals who are as similar as possible to the treatment group. Otherwise we do not have a counterfactual with which to compare the outcomes of the treatment group. Beside difference-in-differences approach, policy evaluation

[17] A necessary condition for difference-in-differences method to work is the parallel trends assumption before the treatment. However, this is not a sufficient condition.

have used for example regression discontinuity design and instrumental variables successfully.

Instrumental variables (IV) have been widely used both in randomized and quasi-experiment evaluation when there is imperfect compliance. The idea behind instrumental variables is to use the variation in the explanatory variable that is caused by the instrument. That is, in the first-stage the explanatory variable is regressed against the instrument. If the instrument is associated with the explanatory variable and does not affect the outcome variable of interest directly (exclusion restriction assumption), in the second stage one utilizes the first-stage's predicted values of the explanatory variable with the controls to get a causal estimate for the explanatory variable.

The commonly used instruments in studies exploring the effect of retirement or reduced working hours include the statutory work-week regulation (Cygan-Rehm and Wunder, 2018; Ahn, 2016) or statutory retirement ages (Kantarci, 2016). Recently, in the health economics literature, the instrumental variables have been accompanied with fixed effects if panel data is available (Cygan-Rehm and Wunder, 2018; Ahn, 2016). Fixed-effects instrumental variable estimation has the advantage that it removes the time-invariant unobserved variables while taking into consideration the endogeneity of the explanatory variable.

The aim of the programme evaluation is to give some recommendations to the policy makers. The object of interest in experiments is often the average treatment effect (ATE). ATE represents how the average outcomes differ in the whole population if we could move everybody from the inactive to active treatment. A some-what more limited effect is the average treatment effect for the treated (ATET). (Abadie and Cattaneo, 2018). In the difference-in-differences setting the effects are average treatment effects for the treated if all study units take-up the treatment. This means that with the example of work hours reduction policies, all individuals need to actually reduce hours worked. However, it is seldom that the effects can be interpreted this way. Many treatments often have only partial compliance to the assignment and so the effects are intention-to-treat (ITT) effects. For some policy evaluation this is enough because the policy maker is interested in knowing whether providing an access to a certain policy is beneficial.

Especially with IV estimates, the effects often lack the external validity and are instead interpreted as local average treatment effects (LATE) (Angrist et al., 1996). LATEs mean the effect on the sub-population of compliers who change their behaviour because of the instrument or the average effect for those who would always be in compliance regardless of the treatment. So these effects cannot be extrapolated to other parts of the population. The researchers should carefully consider what sort of effects they are observing and what kind of policy advice they can draw from those effects.

1.4 Summaries of the essays

1.4.1 ESSAY 1: On the optimal lifetime redistribution and social objectives: a multidimensional approach

In the first essay, which is a joint work with Sanna Tenhunen and Matti Tuomala, we study the optimal redistribution over the life-cycle in an economy where individuals differ with respect to their abilities and their discount rates. Both sources of heterogeneity are unobservable to the government. The heterogeneity in discount rates means that individuals value the future consumption differently. Some individuals are more patient and save for the retirement period while others prefer consuming relatively more at a younger age. The heterogeneity in time preferences towards consumption adds a second dimension to the classical optimal tax problem where individuals only differ with respect to abilities. We study the optimal structure of taxation and focus especially on whether capital income should be taxed or not.

For the social objectives we incorporate a novel type of social welfare function which applies a compromise between the principles of compensation and responsibility much discussed in the equality of opportunity literature. We call this social objective as the Roemer social welfare function. In a two-period model where individuals work in the first period and are retired in the second period, we show that two-dimensional heterogeneity between agents results in optimality of non-zero capital income tax. The analytical results show that the "no distortion at the top" -result with respect to the ability does not hold under the preference heterogeneity. Numerical simulations reveal that the levels of marginal tax rate can change dramatically between different social objectives. From the numerical simulations we also observe that the correlation between preference towards the timing of consumption and ability plays a crucial role.

1.4.2 ESSAY 2: On optimal income taxation when inherited wealth differs

In the second essay, which is a joint work with Matti Tuomala, we study the optimal taxation of capital and labour income when individuals differ in their productivities and initial wealth. Both sources of heterogeneity are unobservable to the government. The initial wealth is exogenous, and in the first period individuals need to decide how much of the wealth is consumed and how much saved. In the second period individuals supply labour effort and consume the returns of savings and income from work. In the first case the government observes the savings from the first period to the second and the labour income and aims to build an efficient tax-and-transfer system which induces individuals to reveal their true type. In this setting we study whether non-linear capital (savings) tax belongs to the optimal tax mix. We also extend the model to include income shifting. In this extension the government does not perfectly observe the labour income or savings but relies on income reporting from the individuals. This induces some individuals to shift income from labour to capital income tax base. Taking this into account, we study the optimal tax mix.

We analytically solve two- and three-type models and numerically solve a four-type model. Our analytical findings, assuming the direction of binding self-selection constraint, suggest that if there are wealth differences and productivity differences between individuals, the optimal tax system includes a non-linear capital income tax. The income shifting also supports including a non-linear capital income tax. In the numerical solutions to the 4-type economy we do not assume *a priori* the binding self-selection constraints. In all of the numerical simulation undertaken there is a role for capital income tax. Numerically we also study the effect of different redistributive preferences, the role of wealth inequality and the correlation between ability and wealth.

1.4.3 ESSAY 3: Top incomes and income dynamics from a gender perspective: Evidence from Finland 1995-2012

In the third essay I study Finnish top incomes. I will first update the earlier analyses of the overall top income shares until the year 2012 using individual income concepts. My main contribution is to study the top incomes from a gender point of view. I use a unique dataset without top-coding and a representative sample of 10 % of the Finnish population. The data cover the years 1995-2012 and I am especially interested in the development measured over time in female representation, mobility and income composition.

I find that women are noticeable under-represented at the top of the income distribution. In the top 10% of the income distribution, excluding realized capital gains, only one in four is a woman and the share has increased only after the financial crisis. Nonetheless, the share of women in the top 1% has improved within the observed 18 years. Also the income composition reveals that at the end part of the observation period women have accrued higher returns from the labour markets as their wage share has increased. The persistence to stay in the top groups is slightly lower for women than for men.

I also study the gender-specific income distributions. The gender-specific income shares reveal that incomes are less dispersed among women. Fitting Pareto-tails to the top of the gender-specific distributions shows that while men's income distribution can be assumed Paretian for the whole observation period, this is not the case for the early observation years for women. For women, the assumption of Pareto tail gets more support after the year 2000. The female top income receivers have caught up with top earning men over time.

1.4.4 ESSAY 4: The effects of working hours reduction on health and labour market exits: Evidence from the Finnish part-time pension program

In the fourth essay I study the effects of working hours reduction on the health outcomes and the labour market exits in the elderly workforce. I focus on the subgroup of individuals who have taken part-time pension during years 1998-2005. The part-time pension program provides an interesting research design because it certainly reduced the hours worked, which is not the case in many voluntary partial retirement schemes. Also the Finnish part-time retirement usually meant that individuals continued in their career jobs so we can confidently infer that the results are not driven by the changes of work place and work community. This is a clear advantage since many countries, for example the US, have different types of jobs offered for the elderly workforce who are searching to reduce their work burden. Also the Finnish part-time pension scheme was very generous, which meant that the disposable income was only slightly affected and so the effects are not driven by income.

I study the question from two points of view. Firstly I study a reform in year 1998 which lowered the eligibility age for part-time retirement. Using a difference-in-differences approach, where I have a treatment group whose eligibility age was 56 and a control group who were eligible at the age of 58, I study the reform effects on purchase of drugs and sickness absence days. These results reveal on average what kind of intention-to-treat effects this reform had in the eligible population. The findings suggest that on average the reform increased the drug purchases but might have had a decreasing effect on the sickness absences, however these results are rather imprecise. The second approach looks at the effects of taking part-time pension in a subgroup of part-time pensioneers. Here I utilize the eligibility ages as instruments and study, with fixed effects instrumental variables estimation, how being on part-time pension affects the health-outcomes and labour market exits. These effects suggest that reducing work hours decreases the drug purchases and early labour market exit risk within the subgroup of compliers, who compared to non-compliers have worse health outcomes before the part-time pension spell. The essay also discusses the robustness of the results and the heterogeneity of the effects.

Bibliography

Aaberge, R. and Atkinson, A. B. (2010). Top Incomes in Norway. In Atkinson, A. B. and Piketty, T., editors, *Top Incomes: Global Perspective*, chapter 9, pages 448–481. Oxford University Press, Oxford.

Aaberge, R., Atkinson, A. B., and Königs, S. (2018). From classes to copulas: wages, capital, and top incomes. *The Journal of Economic Inequality*, 16(2):295–320.

Aaberge, R., Atkinson, A. B., and Modalsli, J. (2013). The ins and outs of top income mobility. Statistics Norway, Research Department, Discussion Papers, No. 762.

Abadie, A. and Cattaneo, M. D. (2018). Econometric Methods for Program Evaluation. *Annual Review of Economics*, 10(1):465–503.

Aghion, P., Akcigit, U., Bergeaud, A., Blundell, R. W., and Hémous, D. (2019). Innovation and Top Income Inequality. *Review of Economic Studies*, Volume 86, Issue 1, January 2019:1–45.

Ahn, T. (2016). Reduction of working time: Does it lead to a healthy lifestyle? *Health Economics*, 25:969–983.

Aiyagari, S. R. (1995). Optimal Capital Income Taxation with Incomplete Markets , Borrowing Constraints , and Constant Discounting. *Journal of Political Economy*, 103(6):1158–1175.

Alstadsæter, A., Jacob, M., Kopczuk, W., and Telle, K. (2016). Accounting for Business Income in Measuring Top Income Shares: Integrated Accrual Approach Using Individual and Firm Data from Norway. Working Paper 22888, National Bureau of Economic Research.

Alvaredo, F., Atkinson, A. B., Piketty, T., and Saez, E. (2013). The Top 1 Percent in International and Historical Perspective. *Journal of Economic Perspectives*, 27(3):3–20.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434):444–455.

Armstrong, M. (1996). Multiproduct nonlinear pricing. *Econometrica*, 64(1):51–75.

Armstrong, M. and Rochet, J.-C. (1999). Multi-dimensional screening : A user's guide. *European Economic Review*, 43:959–979.

Atkinson, A. B. (1997). Bringing Income Distribution in From the Cold. *The Economic Journal*, 107(441):297–321.

Atkinson, A. B., Casarico, A., and Voitchovsky, S. (2018). Top incomes and the gender divide. *Journal of Economic Inequality*, 16:225–256.

Atkinson, A. B. and Piketty, T., editors (2007). *Top Incomes over the Twentieth Century — A Contrast between European and English-Speaking Countries*. Oxford University Press.

Atkinson, A. B. and Piketty, T., editors (2010). *Top Incomes: A Global Perspective*. Oxford University Press.

Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1):3–71.

Atkinson, A. B. and Stiglitz, J. E. (1976). The Design of Tax Structure: Direct versus Indirect Taxation. *Journal of Public Economics*, 6:55–75.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, 344(6186):843–851.

Banks, J. and Diamond, P. (2010). The Base for Direct Taxation. In Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., and Poterba, J., editors, *Dimensions of Tax Design: Mirrlees Review*, chapter 6, pages 548–649. Oxford University Press, Oxford.

Bertrand, M., Black, S. E., Jensen, S., and Lleras-Muney, A. (2019). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labor Market Outcomes in Norway. *The Review of Economic Studies*, 86:191–239.

Bietenbeck, J. and Berniell, I. (2017). The Effect of Working Hours on Health. IZA Discussion Paper Series 10524.

Bivens, J. and Mishel, L. (2013). The Pay of Corporate Executives and Financial Professionals as Evidence of Rents in Top 1 Percent Incomes. *Journal of Economic Perspectives*, 27(3):57–78.

Björklund, A., Roine, J., and Waldenström, D. (2012). Intergenerational top income mobility in Sweden: Capitalist dynasties in the land of equal opportunity? *Journal of Public Economics*, 96(5-6):474–484.

Blau, F. D. and Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3):789–865.

Blomquist, S. and Christiansen, V. (2008). Taxation and Heterogeneous Preferences. *FinanzArchiv / Public Finance Analysis*, 64(2):218–244.

Boadway, R., Marchand, M., Pestieau, P., and Racionero, M. D. M. (2002). Optimal Redistribution with Heterogeneous Preferences for Leisure. *Journal of Public Economic Theory*, 4(4):475–498.

Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Jäntti, M., Eriksson, T., and Österbacka, E. (2007). Nonlinearities in intergenerational earnings mobility: Consequences for cross-country comparisons. *Economic Journal*, 117(519):72–92.

Browning, M. and Meghir, C. (1991). The Effects of Male and Female Labor Supply on Commodity Demands. *Econometrica*, 59(4):925–951.

Chamley, C. (1986). Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives. *Econometrica*, 54(3):607–622.

Choné, B. P. and Laroque, G. (2010). Negative Marginal Tax Rates and Heterogeneity. *American Economic Review*, 100(5):2532–2547.

Christiansen, V. and Tuomala, M. (2008). On taxing capital income with income shifting. *International Tax and Public Finance*, 15:527–545.

Clemens, J., Gottlieb, J. D., Hémous, D., and Olsen, M. (2016). The Spillover Effects of Top Income Inequality. University of British Columbia, mimeo.

Conesa, J. C., Kitao, S., and Krueger, D. (2009). Taxing capital? Not a Bad Idea After All! *The American Economic Review*, 99(1):25–48.

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27(3):79–102.

Corak, M. and Heisz, A. (1999). The Intergenerational Earnings and Income Mobility of Canadian Men : Evidence from Longitudinal Income Tax Data. *Journal of Human Resources*, 34(3):504–533.

Corak, M. and Piraino, P. (2011). The Intergenerational Transmission of Employers. *Journal of Labor Economics*, 29(1):37–68.

Corlett, W. J. and Hague, D. C. (1953). Complementarity and the Excess Burden of Taxation. *The Review of Economic Studies,*, 21(1):21–30.

Cortés, P. and Pan, J. (2019). When Time Binds: Substitutes for Household Production , Returns to Working Long Hours , and the Skilled Gender Wage Gap. *Journal of Labor Economics*, 37(2):351–398.

Crawford, I., Keen, M., and Smith, S. (2010). Value added tax and excises. In Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., and Poterba, J., editors, *Dimensions of tax design: the Mirrlees review*, chapter 4, pages 275–362. Oxford University Press, Oxford.

Cremer, H., De Donder, P., Maldonado, D., and Pestieau, P. (2009). Forced Saving, Redistribution, and Nonlinear Social Security Schemes. *Southern Economic Journal*, 76(1):86–98.

Cremer, H. and Gahvari, F. (1995a). Uncertainty and optimal taxation: In defense of commodity taxes. *Journal of Public Economics*, 56:291–310.

Cremer, H. and Gahvari, F. (1995b). Uncertainty, optimal taxation and direct versus optimal indirect taxation and the tax controversy. *The Economic Journal*, 105(432):1165–1179.

Cremer, H., Pestieau, P., and Rochet, J. C. (2001). Direct versus indirect taxation: The design of the tax structure revisited. *International Economic Review*, 42(3):19.

Cygan-Rehm, K. and Wunder, C. (2018). Do working hours affect health? Evidence from statutory workweek regulations in Germany. *Labour Economics*, 53(October 2017):162–171.

Diamond, P. and Mirrlees, J. A. (1971a). Optimal Taxation and Public Production I: Production Efficiency. *The American Economic Review*, 61(1):8–27.

Diamond, P. and Mirrlees, J. A. (1971b). Optimal Taxation and Public Production II : Tax Rules. *The American Economic Review*, 61(3):261–278.

Diamond, P. A. and Spinnewijn, J. (2011). Capital income taxes with heterogeneous discount rates. *American Economic Journal: Economic Policy*, 3(November):52–76.

Dworkin, R. (1981a). What is Equality? Part 1: Equality of Welfare. *Philosophy & Public Affairs*, 10(3):185–246.

Dworkin, R. (1981b). What is Equality? Part 2: Equality of Resources. *Philosophy & Public Affairs*, 10(4):283–345.

Edlund, B. L. and Kopczuk, W. (2009). Women, Wealth, and Mobility. *American Economic Review*, 99(1):146–178.

Fairfield, T. and Jorratt De Luis, M. (2016). Top Income Shares, Business Profits, and Effective Tax Rates in Contemporary Chile. *Review of Income and Wealth*, 62(August):S120–S144.

Fleurbaey, M. and Maniquet, F. (2018). Optimal Income Taxation Theory and Principles of Fairness. *Journal of Economic Literature*, 56(3):1029–1079.

Fortin, N. M., Bell, B., and Böhm, M. (2017). Top earnings inequality and the gender pay gap : Canada , Sweden , and the United Kingdom. *Labour Economics*, 47(May):107–123.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4):1091–1119.

Golosov, M., Troshkin, M., Tsyvinski, A., and Weinzierl, M. (2013). Preference heterogeneity and optimal capital income taxation. *Journal of Public Economics*, 97:160–175.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2):223–255.

Guvenen, F., Kaplan, G., and Song, J. (2014). The Glass Ceiling and The Paper Floor: Gender Differences among Top Earners , 1981 – 2012. NBER Working Paper No. 20560.

Jacobs, B., Jongen, E. L., and Zoutman, F. T. (2017). Revealed social preferences of Dutch political parties. *Journal of Public Economics*, 156:81–100.

Jacquet, L., Lehmann, E., and Van der Linden, B. (2013). Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory*, 148:1770–1805.

Jäntti, M. and Jenkins, S. P. (2015). Income Mobility. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, pages 807–937. Elsevier Science & Technology.

Jäntti, M., Riihelä, M., Sullström, R., and Tuomala, M. (2010). The trends in top income shares in Finland 1966–2007. In Atkinson, A. B. and Piketty, T., editors, *Top Incomes: Global Perspective*, number 157, chapter 8, pages 1–36. Oxford University Press.

Judd, K. L. (1985). Redistributive Taxation in a Simple Perfect Foresight Model. *Journal of Public Economics*, 28:59–83.

Kanbur, R. and Tuomala, M. (1994). Inherent Inequality and the Optimal Graduation of Marginal Tax Rates. *Scandinavian Journal of Economics*, 96(2):275–282.

Kanbur, R. and Wagstaff, A. (2014). How Useful Is Inequality of Opportunity as a Policy Construct? World Bank Policy Research Working paper 6980.

Kantarci, T. (2016). Effects of Working Part-Time and Full-Time on Physical and Mental Health in Old Age. Tilburg University, Mimeo.

Kaplow, L. (2006). On the undesirability of commodity taxation even when income taxation is not optimal. *Journal of Public Economics*, 90(6-7):1235–1250.

Kleven, J. H., Landais, C., Saez, E., and Schultz, E. (2014). Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark. *The Quarterly Journal of Economics*, 129(1, February 2014):333–378.

Kopczuk, W., Saez, E., and Song, J. (2010). Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937. *The Quarterly Journal of Economics*, 125(1, February 2010):91–128.

Kuznets, S. (1955). Economic growth and income inequality. *American Economic Review*, XLV(1):1–28.

Kuznets, S. and Jenks, E. (1953). Shares of Upper Income Groups in Saving. In *Shares of Upper Income Groups in Income and Savings*, volume I, pages 171–218. NBER.

Lemieux, T., Macleod, W. B., and Parent, D. (2009). Performance Pay and Wage Inequality. *The Quarterly Journal of Economics*, 124(1, February 2009):1–49.

Lockwood, B. B. and Weinzierl, M. (2015). De Gustibus non est Taxandum: Heterogeneity in preferences and optimal redistribution. *Journal of Public Economics*, 124:74–80.

Marchand, M., Pestieau, P., and Racionero, M. (2003). Optimal redistribution when different workers are indistinguishable. *Canadian Journal of Economics*, 36(4):911–922.

Mirrlees, J. (1971). An Exploration in the Theory of Optimal Income Taxation. *The Review of Economic Studies*, 38(2):175–208.

Naito, H. (1999). Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency. *Journal of Public Economics*, 71(2):165–188.

Philippon, T. and Reshef, A. (2012). Wages and Human Capital in the U.S. Financial Industry: 1909-2006. *The Quarterly Journal of Economics*, 127(4, November 2012):1551–1609.

Piketty, T. (2007). Top Incomes Over the Twentieth Century: A Summary of the Main Findings. In Atkinson, A. and Piketty, T., editors, *Top Incomes Over* *the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*, chapter 1, pages 1–18. Oxford University Press, Oxford.

Piketty, T. and Saez, E. (2003). Income Inequality in the United States, 1913-1998. *The Quarterly Journal of Economics*, 118(1, February 2003):1–39.

Piketty, T., Saez, E., and Stantcheva, S. (2014). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy*, 6(1):230–271.

Piketty, T. and Zucman, G. (2014). Capital is Back: Wealth-Income Rations in Rich Countries 1700–2010. *The Quarterly Journal of Economics*, 129(3, August 2014):1255–1310.

Pirttilä, J. and Suoniemi, I. (2014). Public Provision, Commodity Demand, and Hours of Work: An Empirical Analysis. *Scandinavian Journal of Economics*, 116(4):1044–1067.

Ramos, X. and Van de gaer, D. (2016). Approaches To Inequality of Opportunity: Principles, Measures and Evidence. *Journal of Economic Surveys*, 30(5):855–883.

Ramsey, F. (1927). A Contribution to the Theory of Taxation. *The Economic Journal*, 37(145):47–61.

Ravaska, T. (2018). Top incomes and income dynamics from a gender perspective: Evidence from Finland 1995–2012. Labour Institute for Economic Research, Working Papers 321.

Ravaska, T., Tenhunen, S., and Tuomala, M. (2018). On the optimal lifetime redistribution and social objectives: a multidimensional approach. *International Tax and Public Finance*, 25(3):631–653.

Rochet, J.-C. and Choné, P. (1998). Ironing, Sweeping, and Multidimensional Screening. *Econometrica*, 66(4):783–826.

Roemer, J. E. (1998). *Equality of Opportunity*. Harvard University Press.

Roemer, J. E. and Trannoy, A. (2016). Equality of Opportunity: Theory and Measurement. *Journal of Economic Literature*, 54(4):1288–1332.

Roine, J., Boschini, A., and Gunnarsson, K. (2017). Women in Top Incomes: Evidence from Sweden 1974 – 2013. Iza Discussion Paper Series No. 10979.

Roine, J. and Waldenström, D. (2010). Top Incomes in Sweden over the Twentieth Century. In Atkinson, A. B. and Piketty, T., editors, *Top Incomes: Global Perspective*, chapter 7, pages 299–370. Oxford University Press.

Rosen, S. (1981). The Economics of Superstars. *The American Economic Review*, 71(5):845–858.

Sadka, E. (1976). On Income Distribution, Incentive Effects and Optimal Income Taxation. *The Review of Economic Studies*, 43(2):261–268.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *The Review of Economic Studies*, 68(1):205–229.

Saez, E. (2002). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics*, 83:217–230.

Seade, J. K. (1977). On the shape of optimal tax schedules. *Journal of Public Economics*, 7(2):203–235.

Shorrocks, A. (1978). Income inequality and income mobility. *Journal of Economic Theory*, 19(2):376–393.

Slemrod, J. (1996). High-Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response. In Feldstein, M. and Poterba, J., editors, *Empirical Foundations of Household Taxation*, chapter 6, pages 169–192. University of Chicago.

Stern, N. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, 17(2):181–211.

Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics*, 17:213–240.

Stiglitz, J. E. (2018). Pareto efficient taxation and expenditures: Pre- and redistribution. *Journal of Public Economics*, 162(September 2017):101–119.

Suoniemi, I. (2017). Intergenerational mobility and equal opportunity, evidence from Finland. Labour Institute for Economic Research, Working Papers 312.

Tenhunen, S. and Tuomala, M. (2010). On optimal lifetime redistribution policy. *Journal of Public Economic Theory*, 12(1):171–198.

Tuomala, M. (2016). *Optimal Redistributive Taxation*. Oxford University Press.

Tuomala, M. and Tenhunen, S. (2013). On the design of an optimal non-linear tax/pension system with habit formation. *International Tax and Public Finance*, 20:485–512.

Van de gaer, D. (1993). *Equality of opportunity and investment in human capital*. Phd dissertation, Katholieke Universiteit Leuven.

Wolfson, M., Veall, M., and Brooks, N. (2016). Piercing the Veil – Private Corporations and the Income of the Affluent. *Canadian Tax Journal*, 64(1):1–25.

2 ON THE OPTIMAL LIFETIME REDISTRIBUTION AND SOCIAL OBJECTIVES: A MULTIDIMENSIONAL APPROACH

Terhi Ravaska, *University of Tampere, Tampere, Finland*
Sanna Tenhunen, *Finnish Centre for Pensions, Helsinki, Finland*
Matti Tuomala, *University of Tampere, Tampere, Finland*

Abstract

We characterise[1] optimal redistribution policy when there are differences not only in individuals' productivities but also in their tastes towards the timing of consumption, i.e. some are patient and others impatient in consumption over the life-cycle and this preference together with productivity is non-observable to government. We consider different social objectives and incorporate a novel approach taken in the spirit of Roemer (1998) and Van de gaer (1993). This approach applies a compromise between the principle of compensation and the principle of responsibility. We derive analytical expressions which describe the optimal distortion (upward or downward) in saving.

As the multidimensional problems become very complicated, to gain a better understanding, we also numerically examine the properties of an optimal lifetime redistribution policy. We find support for a non-linear tax/pension program in which

[1]Funding from the Strategic Research Council of Academy of Finland, No. 293120 (STN-WIP-consortium) is gratefully acknowledged. Ravaska also gratefully acknowledges the funding from OP Group's Scientific Foundations. We would like to thank two anonymous referees and Sara LaLumia for valuable comments. We also thank the seminar audiences at the 34th Summer Seminar of Finnish Economists, the XXXVII Annual Meeting of the Finnish Economic Association,the 71nd Annual Congress of the International Institute of Public Finance in Dublin and the Nordic workshop on tax policy and public economics in Oslo for the discussion.

impatient types are taxed at the margin and patient low ability types are subsidized in their retirement consumption. Numerical simulations show quite big differences in terms of the levels of marginal tax rates between different social objectives, indicating that the optimal income taxation results are sensitive for the choice of the social planner's goals.

Keywords: Optimal taxation, Lifetime redistribution, Heterogeneous time preferences
JEL classification: H21, H55, D71

2.1 Introduction

The assumption that differences in lifetime earnings can be completely explained by time preferences, is obviously an unrealistic one, just as it would be a simplification of reality to explain them as a result of differences in ability. A more realistic model should take both into account. There are well-known technical difficulties related to incentive constraints to study multi-dimensional optimal tax problems including both of the elements. Another problem is how to incorporate heterogeneous preferences into a social welfare function (SWF) in analysing optimal tax policy. Social welfare functions can be quite straightforwardly parameterized when individuals have identical preferences represented by a utility function. In the case of a one-dimensional population, there are two possible ways to observe differences in economic outcomes. Namely, if people have identical preferences but they differ in abilities, we are back in the Mirrlees (1971) model. In the opposite case, the diversity of preferences is the sole source of inequality.

In the case of diversity in preferences some people would, however, say that if individuals have the same opportunities while their choices may differ, there is no ethical basis for redistributive taxation. According to this view individuals should be compensated for circumstances which they have no control of, such as their family background or disability at birth. On the other hand, individuals should be held responsible for circumstances which they can control, such as how many hours or weeks they work. Hence, no redistribution should take place based on such choices. The former is referred to as the *principle of compensation* and the latter the *principle of responsibility* (see Fleurbaey (1994); Roemer (1998)). By the principle of compensation, it is fair to redistribute from high ability to low ability individuals. By the principle of responsibility, it is unfair to redistribute from the consumption lovers toward the leisure lovers. In the one dimensional population, those principles are easy to apply. For example, if individuals differ only according to their earnings ability (wage rate) and not in their preferences, then the principle of compensation reduces to a maximin criterion whereby the tax and transfer system should provide as much compensation as possible to the worst off people. If individuals differ solely in preferences, the principle of responsibility calls for no redistribution at all because everybody has the same opportunities. It would be unfair to redistribute based on tastes. The standard welfarist approach can obtain this result only in the case where social marginal utilities of net income are the same across individuals (absent transfers).

In a multidimensional world the problem of choosing different utility functions for representing non-identical preferences is more complex. If individuals have different preferences, it isn't clear in which way to weight their utilities in a social welfare function. It can be argued that the fundamental distinction is not so much between earning abilities and preferences but between those factors which are beyond the control of individuals, and those which are purely a matter of individual choice. That means redistribution policies should aim to eliminate disparities in matters beyond individual control, but should be neutral about those matters which are within their control. How to apply these two principles? There is a fundamental conflict between these principles. Namely, even in a world of perfect information with lump sum redistribution tools the government cannot generally satisfy these two principles at the same time.

This paper studies the optimal lifetime redistribution policy within a cohort with heterogeneity in earnings ability and preferences. The heterogeneity in preferences arises because tastes towards the timing of life-cycle consumption differ between individuals; some individuals are more present-oriented in consumption and so save less for the retirement period. These preferences are parametrized with different discount rates[2] meaning that the more

[2]One interpretation for the differences in discount factors beside consumption preferences

present-oriented consumer has a higher discount rate and thus a lower discount factor in his/her utility function. The differences in preferences differ from the myopic consumers because myopic consumers *ex-post* prefer saving more in the earlier periods and this justifies government interventions. However, in our paper there is true variation in tastes towards timing of consumption.

Adding preference heterogeneity raises the difficulty of how to choose the appropriate social welfare function. One way is to assume a paternalistic government as with myopic consumers (Cremer et al. (2009); Tuomala and Tenhunen (2013)). With genuine differences in preferences, there are some recent contributions that incorporate heterogeneous time preferences into optimal tax analysis while remaining agnostic about the appropriate cardinalization (Cremer et al., 2009; Tenhunen and Tuomala, 2010; Tuomala and Tenhunen, 2013). There is also a growing body of literature which studies the multidimensional optimal tax problem by avoiding the technical complications by assuming multidimensionalities can be represented with one-dimensional aggregation of the multidimensional characters (e.g Chone and Laroque (2010); Lockwood and Weinzierl (2015) for labour income taxation). The papers closest to ours are by Golosov et al. (2013) and Diamond and Spinnewijn (2011). The first one studies savings taxation and preference heterogeneity under the assumption that there is perfect correlation between the two dimensions, which effectively makes the problem one-dimensional. Diamond and Spinnewijn (2011) consider a model with jobs and differences in savings preferences. They simplify the analysis by studying linear capital tax. In this paper we consider two-dimensional heterogeneity including non-linear tax instruments in a discrete type setting and the analysis is completed with numerical simulations.

In the spirit of Roemer (1998) and Van de gaer (1993)[3]our approach applies a compromise between the principle of compensation and the principle of responsibility. For individuals with the same discount rates but different wage

---

could be that there are individuals who expect to live shorter lives and therefore emphasize the first period consumption. Fleurbaey et al. (2014) study these kind of longevity differences and redistribution.

[3]Bossert (1995) and Fleurbaey (1994) have studied the idea of compensating inequalities due to circumstances only, while leaving other inequalities untouched.

rates, the maximin criterion is applied. Thus we have a social ordering over each discount rate group. Then we aggregate over discount groups so that the minimum utility levels for different discount groups are averaged. The least well off of each preference groups are added together. In other words, a zero aversion of inequality can be applied along the dimension of responsibility (in our case time preference) whereas a high aversion to inequality is acceptable along the dimension of circumstances (in our case skill)[4].

Our model consists of two periods, where individuals work only during the first period and decide how much to save for the second period. Our focus is on the distortions in savings decisions. The paper continues the research done by Tenhunen and Tuomala (2010) and Tuomala and Tenhunen (2013)[5] by introducing the Roemer social welfare function (RSWF) which, to our knowledge, has not been studied previously. This representation of the social goals takes into account the principles of compensation and responsibility as noted above. Since the aim is to model an economy with multidimensional heterogeneity, the analytical results no longer reveal the signs of the distortions. For this reason numerical methods are used. Our results can also be interpreted in absence of private savings. In this case the second period consumption is publicly provided pension and thus we can extend our analysis into studying the optimal retirement plans in our model economy.

The contribution of this paper is to study optimal lifetime redistribution problem under heterogeneous preferences and introduce the RSWF as social goals. The main finding is that in the full 4-type model the saving decisions of the patient low-ability type and the impatient high-ability type are distorted at the margin. The numerical simulations show that the size of the distortions depend upon the correlation between ability and preferences. The impatient high-ability type has a positive marginal tax on savings. The patient low-ability type has a negative marginal tax (subsidy) when the correlation

[4]Fleurbaey and Maniquet (2006) advocate a social welfare function based on fairness principles that puts a greater weight to "working poors" if preferences differ towards leisure. In an intertemporal model like ours this could mean that the more patient poor should have a greater weight. However, as discussed, we are more agnostic about this normative dimension and instead use the Roemer social welfare function as described in the next paragraph.

[5]Tenhunen and Tuomala (2010) studied optimal life-time redistribution in 4-types setting where government's objective is either utilitarian or paternalistic and consumer preferences are approximated with Cobb-Douglas utility function. Tuomala and Tenhunen (2013) studied how habit formation affects the optimal tax and pension scheme under heterogeneous preferences.

between ability and preferences is below 0.5. For greater correlations the impatient low-ability type faces a positive marginal tax while the patient low-ability type has a zero marginal tax. Our results also show that governments with more redistributive goals need to make sure that the patient high-ability individuals do not mimic the low-ability type while for governments with utilitarian social objectives this incentive-compatibility constraint is not binding.

The structure of the paper is the following. In the first section we present the benchmark model where the time preference and ability are perfectly correlated. Then in section 2.3 we extend the model to include three types, first by pooling the low-ability types into one time preference group and in another case by pooling the high-ability types into one time preference group. In section 2.4 we include all the four types in the model. Analytically the direction of distortions cannot be determined so in the end of each section we show with numerical simulations what kind of distortions occur for each type. Section 2.5 concludes.

2.2 A benchmark model

2.2.1 Two types with a positive correlation between skill and discount factor

Unlike in the original Mirrlees model, we assume that individuals differ not only in productivity but also in time preference[6]. As a benchmark we use a simple two-type model, similar to the much used two-type model first introduced by Stern (1982) and Stiglitz (1982). Each individual has a skill level reflecting his/her wage rate, denoted by n, and differences in time preferences are represented by a discount factor, denoted δ. We denote low skill and low discount factor by the superscript L and high skill and high discount factor by the superscript H. The assumption of positive correlation implies that $\delta^L < \delta^H$. The proportion of individuals of type i in the population is N^i,

[6]Sandmo (1993) considers a case where people differ in preferences, but are endowed with the same resources. Tarkiainen and Tuomala (1999, 2007) also consider a continuum of taxpayers simultaneously distributed by skill and preferences for leisure and income.

with $\sum N^i = 1$.

As is well known, due to the results of Atkinson & Stiglitz (1976), under a mild separability assumption income taxation does not need be supplemented by other taxes. Saez (2002) argues that the Atkinson-Stiglitz result of commodity taxes holds only when each individual has identical discount rates. He argues that individuals with higher earnings save relatively more which suggests that high-skilled individuals are more likely to have higher discount factors and thus there is a role for taxing savings beside labour taxation. In this case, the discount factor is positively correlated with productivity level. Diamond and Spinnewijn (2011) study heterogeneous discount factors and capital taxation and show that perfect correlation between ability and discount factor isn't necessary required in order to derive the result of positive capital taxation for the high-skilled type. However, this result is more robust if there is positive correlation between skill and discount preference. We also assume in our benchmark model that the high-ability individuals have higher discount factors.

Thus, we take as a starting point a separable utility with positive correlation between discount factor and productivity.[7] This assumption is relaxed in the subsequent sections.

The life-time utility of an individual of type i is additive in the following way:

$$U^i = u(c^i) + \delta^i v(x^i) + \psi(1 - y^i), \tag{2.1}$$

where c and x denote consumption when young and old respectively and y is labour supply when young[8]. Utility function is increasing in c and x and decreasing in y and it is strictly concave, i.e. $u', v', \psi' > 0$, $u'', v'', \psi'' < 0$. We also assume that all goods are normal. Without government intervention, in-

[7]Alternatively the same outcome could be reached by assuming homothetic preferences and linear Engel curves.

[8]Another specification for utility function could be $U^i = u(c^i)/\delta^i + v(x^i) + \psi(1 - y^i)$ which would imply that the ones with higher savings rate are less willing to increase work for additional money. We justify our choice for the utility representation as we are considering lifetime redistribution where the timing of the retirement consumption is a far-off event instead of a nearby event, and in this kind of setting the representation in the text is standard. We also want to compare our results to earlier studies, which are done with this specification. It is obvious that a different choice of utility representation will engender a different optimal solution and affect our results. Effectively the problem is reversed and different IC constraints are binding than in the current setting. See also discussion in Diamond and Spinnewijn (2011).

dividuals choose their consumption level for both periods and labour supply by maximizing the utility function (1) subject to their budget constraint.

To introduce returns to capital and the possible taxation thereof, it is necessary to consider a two-period model. Individuals are free to divide their first period income between consumption, c, and savings, s. Each unit of savings yields an additional $1+\theta$ units of consumption in the second period after-tax income, x. As a further simplification we assume that the return on savings, θ, is fixed, which may be justified by assuming that we consider a small open economy facing a world capital market. Consumption in each period is given by $c^i = n^i y^i - T(n^i y^i) - s^i$ and $x^i = (1+\theta)s^i$, $i = L, H$.

The government wishes to design a lifetime tax system that may redistribute income between individuals in the same cohort. There is asymmetric information in the sense that the tax authority is informed neither about individual skill levels, labour supply nor discount rates. It can only observe before-tax income, ny. In this setting, where tax on both earnings and savings income are available, we examine whether or not savings ought to be taxed. The separability assumption makes it possible to isolate the significance of variations in time preferences.

In the government's problem the control variables are c^i, x^i and y^i. If we assume that there are no private savings, we have a model of labour income taxation in the first period and public provision of pension in the second period. In a case with perfect correlation between skills and discount factors, Roemer's and Van de Gear's approaches are equivalent with the maximin social welfare function. Here the government maximizes the welfare of the worst-off group:

$$U^L = u(c^L) + \delta^L v(x^L) + \psi(1 - y^L), \tag{2.2}$$

subject to the revenue constraint

$$\sum N^i (n^i y^i - c^i - r x^i) = R, \tag{2.3}$$

where $r = \frac{1}{1+\theta}$ and the self-selection (i.e. incentive-compatibility) constraint[9]

$$u(c^H) + \delta^H v(x^H) + \psi(1 - y^H) \geq u(c^L) + \delta^H v(x^L) + \psi(1 - \frac{n^L}{n^H} y^L). \quad (2.4)$$

Multipliers λ and μ are attached to the budget constraint and the self-selection constraint, respectively. The Lagrange function of the optimization problem and the first-order conditions are presented in the appendix A.

Our main interest is in the marginal taxation of savings[10]. For this purpose the first-order conditions are written in the form $(\frac{u_c}{v_x})^i = \frac{\delta^i}{r}(1 - d^i)$, where the left hand side is individual i's marginal rate of substitution between consumption in the first and in the second period and d^i is the distortion. A positive (negative) d^i implies that type i should have an implicit tax (subsidy) on savings. It is useful to define a relative difference in discount factors as $\Delta^{ij} \equiv \frac{\delta^i - \delta^j}{\delta^j}$ for any pair of discount factors. The first-order conditions (presented in Appendix A equations A.2-A.7) imply that

$$\begin{aligned} d^L &= (\varphi^1 - 1)\Delta^{HL} \\ d^H &= 0, \end{aligned} \quad (2.5)$$

where $\varphi^1 = \frac{N^L}{N^L - \mu^{HL}}$. The returns to savings of type i should not be taxed when d^i is zero. As $d^H = 0$, the optimal implicit marginal tax rate for the high-skill type is zero. When we assume, empirically plausibly, that $\delta^H > \delta^L$, we have $d^L > 0$ implying implicit taxation of savings for the low-skilled type. This is the same result as in utilitarian case by Diamond (2003).

As a result of the two-dimensional heterogeneity, a tax on capital income is an effective way to relax an otherwise binding self-selection constraint. This is

[9]The direction of the binding self-selection constraint is assumed to be, following the tradition in the one-dimensional two-type model, from high-skilled individual towards low-skilled individual. This pattern is also confirmed in the numerical simulations.

[10]In the numerical solution we also consider the marginal labour income tax rates. As has become conventional in the literature we may interpret the marginal rate of substitution between gross and net income as one minus the marginal income tax, $\frac{\psi'(\frac{ny}{n})}{nu_c} = 1 - T'(ny)$, which would be equivalent to the characterization of the labour supply of an agent facing an income tax function $T'(ny)$. As in our model the heterogeneity shows up in the discount factor of the second period instant utility, the analytical results do not differ for the optimal labour income tax for the two distinct preference groups. The marginal labour income tax rates satisfy the usual properties; $T(n^L y^L) > 0$ and $T(n^H y^H) = 0$.

Table 2.1 Parameterization.

Fraction of individuals in each group	$N^i = 0.5$ for $i = L, H$
Discount factor	$\delta^L = 0.6, \delta^H = 0.8, r = 0.95$
Productivities (wages)	$n^L = 2, n^H = 3$

because even under separability the mimicker and the individual mimicked do not save the same amount. A high-skilled individual choosing to mimic the low-skilled type values savings more than a low-skilled individual, since discounting of the future is less for the patient high-ability mimicker. Thus, taxing savings relaxes the self-selection constraint. Or put in another way: distortions generate second-order efficiency costs but first-order redistributional benefits.

2.2.2 Numerical simulations

Adding multidimensionality to a constrained optimization problem causes that the directions of the distortions cannot be determined from the analytical results. To gain information on the properties of the optimal redistribution policy we rely on numerical simulations[11]. These simulations reveal the binding incentive-compatibility constraints and so we can determine the allocations that make the agents reveal their true characteristics.

In the numerical examples we assume the following separable form of utility function: $U^i = -\frac{1}{c^i} - \delta^i \frac{1}{x^i} - \frac{1}{1-y^i}$ (CES) and parameterize the model as shown in table 2.1. No a priori assumptions of the binding self-selection constraints are made in the numerical simulations.

In the benchmark model, with perfect positive correlation between the skill and time preferences, numerical simulation verifies the assumption that the only binding self-selection constraint is type H considering mimicking type L[12]. Table 2.2 presents the optimal consumption, labour and utility levels, the marginal tax rates for savings and labour and the replacement rate (x/ny). The additional results are presented in appendix B tables B1 and B2.

[11] The numerical procedure is described in Tenhunen and Tuomala (2010) Appendix B.

[12] The slackness of the other self-selection constraints is also checked by calculating the difference in utilities when mimicking and when not.

Table 2.2 Consumption in period 1 and 2 (c and x), labour supply (y), utility level (U), marginal tax rates on labour income and savings (T' and d), replacement rate (x/ny).

maximin	c	x	y	U	T'	d	x/ny
Type L	0.67	0.41	0.38	-4.58	29.43	41.71	53.07
Type H	0.75	0.69	0.57	-4.79	0	0	40.62
utilitarian							
Type L	0.67	0.51	0.50	-4.67	6.30	6.72	51.20
Type H	0.79	0.72	0.54	-4.56	0	0	44.40

The results presented in table 2.2 show that in an optimal solution the replacement rate decreases in earnings, meaning that the optimal pension system is progressive. There is a positive distortion i.e. marginal saving tax for the low-productivity type[13]. Compared to the case where there are no differences in preferences (and all have a high discount factor) the replacement rate in the current case is much lower for the low-productivity worker. This is because the level of the second period consumption is much smaller due to the lower discount factor.

The heterogeneity in preferences impacts the optimal distortions significantly as without the difference in the discount factor the optimal distortion is zero irrespective of the social welfare function. As the labour supply y could also be interpreted as the length of career, it means that the different discount factors lead to much shorter careers for the low-productivity worker compared to the case with the same discount factor. The high-productivity type is relatively less affected from introducing the differences in preferences to the economy.

Compared to the utilitarian social welfare function (weighted sum of low and high types' utility functions) the numerical simulations show that the levels of marginal tax rates differ significantly between the two types of objective functions. The average net taxes (shown in Appendix B) also show that there is a great difference between the maximin and utilitarian cases. In the maximin case, the government aims to improve the wellbeing of the low-productivity type by significant subsidies and by distorting the labour supply

[13]In the case of perfect negative correlation, numerical simulations show that L's distortion is a marginal subsidy. See figure 2.1 in section 4 for results with varying correlation.

Table 2.3 Types of individuals.

	low-skilled, n^L	high-skilled, n^H
low delta, δ^L	Type 1	Type 3
high delta, δ^H	Type 2	Type 4

and savings decision relatively more than in the utilitarian case. Also if we interpret y as a length of career or retirement age, we notice that in the maximin case the length of career is much shorter for the low-productivity worker (reduces by approximately 24 % from utilitarian case) but the labour supply decision for high-productivity worker is much less affected by changing the social preferences to maximin (labour supply increases by approximately 5%).

2.3 A three-type case

In this and the next sections we generalise the previous model by giving up the assumption that productivity and time preferences are perfectly correlated. In general, there are now four types of individuals who differ both in productivity and time preferences numbered as in table 2.3[14].

2.3.1 Low-ability types have same time preference

To maintain the tractability, we first simplify the model further by assuming that there are actually only three types. First we explore the case where the low-productivity types all have a low discount factor, δ^L, and are indexed as type 1. It can be justifiable to think that the low-productivity types are pooled together either because they have homogeneous preferences towards savings or because their saving ability is constrained i.e. heavier time discounting is forced. High-productivity types with a low discount factor are denoted as type 3 and high-productivity workers with a high discount factor as type 4. We also assume that utility is given by Eq. (2.1), so we have the the same

[14] In the case of one-dimensional heterogeneity types are ordered usually with respect to their income, consumption or utilities but in a two-dimensional world the ordering is not self-evidently clear.

additively separable form as in the two-types case.

With more than two types there are several possibilities for mimicking. Which of the self-selection constraints bind depends on the interaction between individual preferences and the distributional preferences of the government, which hinge on the time preferences and skill level. A priori no binding constraints are forced but they are determined in the numerical simulation. To shorten the notation, the analytical results are shown only with the binding constraints.

When the government has maximin goals, the government's problem reduces to maximising the welfare of the low-ability type as

$$N^1[u(c^1)+\delta^L v(x^1)+\psi(1-y^1)] \tag{2.6}$$

subject to budget constraint and, without any assumptions of the mimicking behaviour, there are six possible self-selection constraints given by $u(c^i)+\delta^i v(x^i)+\psi(1-y^i) \geq u(c^j)+\delta^i v(x^j)+\psi(1-\frac{n^j}{n^i}y^j)$ for $i,j = 1,3,4$ and $i \neq j$. As before, to solve the distortions for savings we rewrite the first-order conditions (presented in appendix A equations A.9-A.17) in form $\frac{u_c^i}{v_x^i} = \frac{\delta^i}{r}(1-d^i)$, $i = 1,3,4$, where the distortions are given by (exploiting the binding incentive-compatibility constraints from numerical solution)

$$\begin{aligned} d^1 &= 0 \\ d^3 &= \frac{\mu^{43}}{\mu^{31}-\mu^{43}}\Delta^{HL} \\ d^4 &= 0, \end{aligned} \tag{2.7}$$

where $\Delta^{HL} = \frac{\delta^H-\delta^L}{\delta^L} > 0$. Equation 7 implies that the saving decision of type 3, high-productivity impatient individual, is now distorted and thus the "no distortion at the top" (with respect to skill) result holds no more. This results from the fact that in a model with given binding self-selection constraints (4,3) and (3,1) a distortion on savings decision helps to mitigate otherwise binding self-selection constraints. Here the low-skilled type 1 is left undistorted. This implies the following proposition:

Proposition 1

When time preferences differ only among the high-ability type and there is only downward binding incentive-compatibility constraints, the savings decision of type 1 and type 4 are not distorted and hence not taxed at the margin. Type 3 faces a distortion at the margin which can be positive or negative.

The sign of the distortion cannot be determined from the analytical results but numerical solution provides the binding incentive-compatibility constraints and their levels. The solution for numerical simulation[15] is given in table 2.4 and additional results are presented in appendix B table B3. The effects of changing the parameter values are shown in subsection 2.3.3.

Numerical solutions show that the analytically ambiguous sign of the distortion d^3 is positive, implying a tax at the margin and the replacement rate imply a progressive pension scheme. The distortion for the impatient high-productivity worker is helping to relax the incentive for patient high-productivity worker to mimic the impatient one. In this 3-type model the impatient high-productivity worker is relatively better off even though his/her replacement rate is much lower than for the other types. This occurs as (s)he consumes more during the first period as (s)he prefers. Compared to the model where all agents have identical preferences about the timing of the consumption, including impatient workers in the model increases the labour supply of low-productivity workers. The numerical solutions also reveal that the intuition from one-dimensional world does not extend to two-dimensional case since the type 3 has a higher utility than type 4 even though the constraint (4,3) is binding.

In the utilitarian case (analytical results presented in appendix in the equations A.19-28 and additional numerical results in table B5) the main difference from the maximin model is that the labour supply or the length of career are closer between types. Intuitively in the model, where more weight is given to the low-skilled type, this type is better off with the cost of high-skilled type. More importantly the numerical simulations show that the differences between the two cases in the levels of marginal tax rates are significant. This is also seen in the average net tax rates (shown in Appendix B tables B3 & B5).

[15] Due to solvability problem the results with CES utility functions are given with parametric values $N^1 = 0.5, N^3 = 0.254$ and $N^4 = 0.246$.

Table 2.4 Consumption in period 1 and 2, labour supply, utility level, marginal tax rates on labour income and savings, replacement rates; low-skilled grouped together.

maximin	c	x	y	U	T'	d	x/ny
Type 1	0.59	0.47	0.41	-4.67	37.07	0	57.44
Type 3	0.83	0.56	0.52	-4.35	0	27.38	36.05
Type 4	0.77	0.70	0.56	-4.71	0	0	41.92
utilitarian							
Type 1	0.65	0.51	0.50	-4.73	8.88	0	51.00
Type 3	0.83	0.64	0.52	-4.22	0	7.73	41.26
Type 4	0.80	0.73	0.54	-4.53	0	0	44.97

2.3.2 High-ability types have same time preference

Alternatively, suppose the kind of 3-type model where the high-productivity individuals have the same time preference. Now type 1 has individuals with low discount factor and low productivity, type 2 include individuals with high discount factor and low productivity and type 4 includes all high-productivity individuals with high discount factor. Here the maximin social welfare function cannot be applied as it is not self-evident which group is the worst off and thus we introduce for the first time the Roemer social welfare function as discussed in the introduction. The government maximises the Roemer social welfare objective function

$$N^1[u(c^1) + \delta^L v(x^1) + \psi(1 - y^1)] + N^2[u(c^2) + \delta^H v(x^2) + \psi(1 - y^2)] \quad (2.8)$$

subject to the same budget constraint and self-selection constraints as earlier. The first-order conditions are presented in appendix A (equations A.30-38). In this case the distortions are (exploiting the binding incentive-compatibility constraints from numerical solution)

$$\begin{aligned} d^1 &= \frac{\mu^{21} + \mu^{41}}{N^1 - \mu^{21} - \mu^{41}} \Delta^{HL} \\ d^2 &= d^4 = 0 \end{aligned} \quad (2.9)$$

Table 2.5 Consumption in period 1 and 2, labour supply, utility level, marginal tax rates on labour income and savings, replacement rates; high-skilled grouped together.

RSWF	c	x	y	U	T'	d	x/ny
Type 1	0.68	0.43	0.41	-4.57	24.73	37.08	52.6
Type 2	0.56	0.51	0.41	-5.04	43.3	0	63.2
Type 4	0.76	0.7	0.56	-4.73	0	0	41.7
utilitarian							
Type 1	0.67	0.52	0.51	-4.68	3.7	5.32	51.0
Type 2	0.62	0.57	0.51	-5.07	12.4	0	55.7
Type 4	0.79	0.73	0.54	-4.55	0	0	44.7

Now only the saving decision of the impatient low-productivity type is distorted.

Proposition 2

When time preferences differ only among the low-ability type and there is only downward binding incentive-compatibility constraints, the savings decision of type 2 and type 4 are not distorted and hence not taxed at the margin. The saving decision of type 1 is distorted at the margin. This distortion can be positive or negative.

The binding incentive-compatibility constraints are (2,1), (4,1) and (4,2). Table 2.5 presents the results with the RSWF and utilitarian social welfare function (additional results are found in appendix B table B.4 and B.6 and sensitivity analysis with different parameter values are shown in the next subsection).

The numerical solutions reveal that type 1's distortion is positive, so there is a tax at the margin and this distortion is quite significant. However, when comparing this distortion to the case of perfect correlation between preferences and productivities, the distortion is not as large. It seems that introducing some patient low-productivity workers to the model facilitate to soften the distortion for the impatient ones in a significant way.

In the RSWF case it is also noticeable that in the optimal solution type 1 and 2 work the same amount, thus the different time preference only leads to different divisions of consumption between periods. The same pattern applies also to the utilitarian case. Here the pattern of replacement rates implies again a progressive pension system, and also indicates that in the optimum the government distorts the patient low-productivity type's work incentives significantly. As before, the most striking difference between the different social goals are in terms of the levels of distortions and the length of the career or retirement age thereof. The average tax rates (shown in Appendix B) also support this difference. In utilitarian case the labour supply is closer to similar between types.

We can conclude from both of the 3-type cases that the saving decision of the less patient individual are distorted in margin. In the case where high-productivity types have the same time preference, the low-productivity types' labour decisions are the same but due to the differences in their time preferences for consumption, the overall utility differs.

2.3.3 Comparative statistics

In order to see what kind of effects the parameters have for the numerical results, in this section we let them vary. Table 2.6 presents the case for varying discount factor in the case of pooling the low-productivity workers. In benchmark those are set to 0.6 and 0.8 respectively for low- and high-productivity type. The distortion for saving is getting smaller as the discount factor is getting closer to the higher discount factor. In the absence of private savings the replacement rates are increasing for types 1 and 3 the closer we get to the higher discount factor. When the discount factors are the same for low-productivity and high-productivity workers, there is no distortion for saving's decision. The large distortions in the RSWF case can be thus accounted for by differences in time preferences. Also the simulation suggests that 0.6 is a lower bound with respect to the time preference for our model, as below that the solver either does not find a solution or there are no binding constraints.

Another case is to vary the sizes of the different types. Table 2.7 present these comparisons. The first thing to notice is that in the RSWF optimization a small size of the low-productivity workers means relatively more redistribution to their end. The replacement rates for high-productivity types are nearly the same in every case. The large marginal labour tax rates make sure that the high-productivity types do not mimic the low-productivity type. The

Table 2.6 Comparison of the solutions with different discount rates, low-ability types pooled together.

δ^L	0.6			0.65			0.7			0.75		
Type	1	3	4	1	3	4	1	3	4	1	3	4
U	-4.7	-4.4	-4.7	-4.8	-4.4	-4.7	-4.9	-4.5	-4.7	-4.9	-4.6	-4.7
T'	37	0	0	38	0	0	38	0	0	38	0	0
d	0	27	0	0	19	0	0	13	0	0	6	0
x/ny	57	36	42	59	38	42	60	39	42	61	41	43

Table 2.7 Comparison of the solutions with varying group sizes, low-ability types pooled.

N^i	0.2	0.4	0.4	0.25	0.375	0.375	0.3	0.35	0.35
Type	1	3	4	1	3	4	1	3	4
U	-4.3	-4.3	-4.6	-4.4	-4.3	-4.6	-4.5	-4.3	-4.7
T'	48	0	0	40	0	0	38	0	0
d	0	28	0	0	28	0	0	28	0
x/ny	137	37	43	100	37	43	81	37	43
N^i	0.35	0.325	0.325	0.4	0.29	0.31	0.45	0.295	0.255
Type	1	3	4	1	3	4	1	3	4
U	-4.5	-4.3	-4.7	-4.6	-4.3	-4.7	-4.6	-4.3	-4.7
T'	35	0	0	32	0	0	30	0	0
d	0	28	0	0	26	0	0	24	0
x/ny	72	36	42	65	36	42	60	36	43

distortion for savings is not much affected by varying the sizes of the groups.

In the second case, where the high-productivity types are pooled together, solving the maximization problem with different discount factors is much more complex.[16] Table 2.8 presents the results which show that the replacement rates are increasing for type 1 when his/her patience increases but this does not significantly affect other types' replacement rates. Also the size of the discount factor seems to have quite significant effect on the level of the marginal savings tax.

[16]In fact, with the CES function, in order to see what kind of effects the discount rates have for the results, the size of the groups needs to be modified: the sizes of the groups in this exercise are set to $N^1 = 0.2$, $N^2 = 0.3$ and $N^3 = 0.5$.

Table 2.8 Comparison of the solutions with varying discount factors, high-ability types pooled.

δ^L	0.6			0.65			0.7		
Type	1	2	4	1	2	4	1	2	4
U	-4.6	-5	-4.7	-4.7	-5	4.7	-4.8	-5	-4.7
T'	24	42	0	28	41	0	31	40	0
d	36	0	0	29	0	0	20	0	0
x/ny	53	63	42	55	63	43	57	62	42

2.4 Roemer social welfare function and 4 types

Finally we include all four types in our model. Now in the economy there are both impatient and patient high-productivity workers as well as impatient and patient low-productivity workers, labelled as in table 2.3. In order to study optimal taxation in the RSWF framework we make a compromise between the principles of compensation and responsibility by firstly computing the minimum within each responsibility group (discount rates here) and then applying the utilitarian criterion. This means that for individuals with the same discount rates but different wage rates the maximin criterion is applied and thus we have a social ordering over each discount group. Then these minimum numbers are added together[17]. Now we have

$$N^1[u(c^1) + \delta^L v(x^1) + \psi(1 - y^1)] + N^2[u(c^2) + \delta^H v(x^2) + \psi(1 - y^2)]. \quad (2.10)$$

The government maximizes now Eq. 2.10 subject to self-selection constraints (without any assumptions of the mimicking behaviour there are twelve possible self-selection constraints) given by
$u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \geq u(c^j) + \delta^i v(x^j) + \psi(1 - \frac{n^j}{n^i} y^j)$ for $i, j = 1, .., 4$ and $i \neq j$, and the budget constraint $\sum N^i(n^i y^i - c^i - rx^i) - R \geq 0$.

Specifications with CES function in this 4-type case turn out to be harder to solve numerically than the earlier cases[18]. The binding constraints ((3,1), (3,2), (4,2) and (4,3)) are exploited in deriving the first-order conditions (Appendix

[17] Or we can first calculate the average utility in each skill group and then apply the maximin criterion to such average figures.

[18] For example there is no binding constraints in the zero correlation case.

A equations A.39-51) and distortions. The distortions are

$$
\begin{aligned}
d^1 &= 0 \\
d^2 &= \frac{\mu^{32}}{N^2 - \mu^{32} - \mu^{42}} \Delta^{LH} \\
d^3 &= \frac{\mu^{43}}{\mu^{32} + \mu^{31} - \mu^{43}} \Delta^{HL} \\
d^4 &= 0.
\end{aligned}
\tag{2.11}
$$

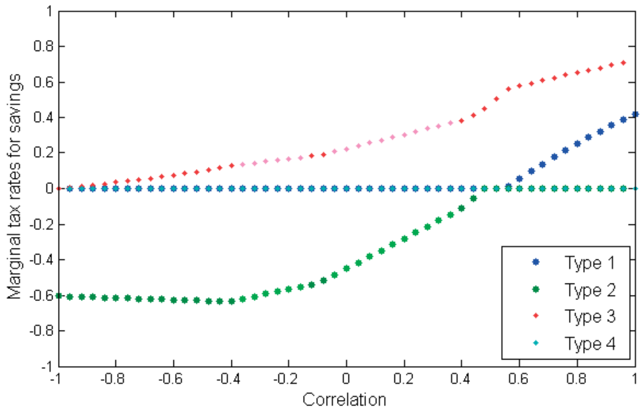
Proposition 3

When individuals differ in both productivity and time preferences and there are only downward binding incentive-compatibility constraints, then savings decisions for type 2 and type 3 are distorted at the margin. The distortions can be negative or positive. There is no distortion for the types 1 and 4. The size and direction of the distortion depends heavily on the correlation between skills and preferences. The "no distortion at the top" (w.r.t ability) result does not hold anymore.

Figure 2.1 shows the simulation results with varying correlation between productivity and preferences. From the simulation results we notice that type 3 has a positive marginal tax. Type 2 has a negative marginal tax when the correlation between productivity and time preferences is below 0.5. Also an interesting feature occurs when there is strong correlation between preferences and productivity. Here the type 1 faces a positive tax but the patient type is not distorted.

From figure 2.1 we see that the problematic points in numerical simulations arise when we come closer to the zero correlation (i.e. group sizes are equal for all 4 types). This is intuitive because moving to either end of the correlation line we come closer to the two-type kind of economy which is generally easier to solve. Studying the optimal bundles in the case that is as close as possible to zero correlation we can conclude that interestingly the labour supplies of impatient and patient low-productivity workers are nearly the same. In the RSWF case this occurs because the saving decision of the patient individual is heavily subsidized and this leads to the situation that type 2's replacement rate is also high. Introducing the impatient high-productivity worker to the

Figure 2.1 Marginal tax on saving and correlation between time preference and ability. Linearly interpolated values are drawn with lighter colour.



model leads to significantly higher marginal tax rates on labour compared to the three-type model. Interpreting the results without private savings, it can be noticed that the pension system is progressive, i.e. the replacement rates are smaller for high-productivity workers.

In the case of utilitarian social welfare function the binding self-selection constraints are (3,1), (3,2) and (4,3). With our parametrization it seems that only when the social goals are more redistributive does the government need to take care that the patient high-ability individual does not mimic patient low-ability type. Compared to the RSWF case, the distortions are significantly smaller and the labour supply of the low-productivity type is greater, which is in line with the earlier results.

Figure 2.1 also provides information of the robustness of the earlier results with different distribution of types. The lighter values are linearly interpolated values as the program cannot solve the optimization problem in certain cases. In the figure we can see that the savings decision of type 2 is heavily subsidised when there is stronger negative correlation between time preference and ability. On the other end of the correlation line, type 3's saving decision is heavily distorted downward at the margin. Type 4 is undistorted in every correlation. Compared to Tenhunen and Tuomala (2010) which studies the similar kind of economy but with utilitarian government, the shape of the curve is similar but here the absolute values of marginal taxes and subsidies are greater. Also, here the taxes for type 1 become non-zero with slightly smaller correlations. To summarize briefly the results about the marginal labour income tax rates we can say that for different correlations both low-skilled types are taxed at the margin in the optimal result. The levels of these taxes are relatively high and stable in different correlations points.

2.5 Conclusion

In this paper we have continued and extended the work of Tenhunen & Tuomala (2010) and Tuomala & Tenhunen (2013) by introducing the Roemer social welfare function. Instead of solely examining an economy where the government is utilitarian, we have considered an economy where the social preferences aim to maximize the welfare of those who have lower productivity while respecting the individuals' true discount rates for future consumption. The multidimensionality stemming from the differences in preferences and productivities makes it unfeasible to fully satisfy these principles of responsibility and compensation at the same time but we have offered one way to derive a compromise of the principles into the social welfare maximization problem.

In the context of a two-period model, we mainly studied the savings distortions and the optimal redistribution policy within a cohort. The results are derived analytically and numerically in several three- and four-types settings. The numerical simulations are done without simplifying the problem to one-dimensional. Also with numerical simulations we do not need to make a priori assumptions about the binding incentive-compatibility constraints. The three-type settings are somewhat easier to solve compared to the four-type models and they can also be used as an insight for the robustness of the four-types results. The numerical solutions help to reveal the sign of the distortion (upwards or downwards) in savings and labour supply behavior compared to the first-best case. The numerical results are compared to the utilitarian one to determine how the objective function affects the results. We have implicitly assumed that the government can commit to a lifetime tax in order to carry out the optimal redistribution policy.

In the lifetime context of labour supply we find that retirement age (length of career) is much lower (shorter) in the RSWF case. The results also indicate that irrespective of the goals of the government the pension system is progressive, i.e. the replacement rates decrease with income. We have considered a case where the differences in preferences are true in a sense that the social planner has no reason to correct for these differences. We learn that in this setting asymmetric information and heterogeneity lead to the savings distortion for impatient individuals at the margin. We have also relaxed the common assumption of positive correlation between skills and time preferences. The correlation between skills and preferences is in an important role. At the margin the patient low-type is subsidized until the correlation gets large and then the low-ability impatient type is taxed at the margin.

The final lesson from our work is that the different goals of the government in welfare maximization cause large differences in the implicit distortions and labour market outcomes. This demonstrates that the government's goals do not only have secondary effects on the optimal taxation results but have an important role in consideration of the levels of distortions.

Bibliography

Bossert, W. (1995). Redistribution mechanisms based on individual characteristics. *Mathematical Social Sciences*, 29:1–17.

Chone, P. and Laroque, G. (2010). Negative marginal tax rates and heterogeneity. *American Economic Review*, 100(5):2532–47.

Cremer, H., De Donder, P., Maldonado, D., and Pestieau, P. (2009). Forced Saving, Redistribution, and Nonlinear Social Security Schemes. *Southern Economic Journal*, 76(1):86–98.

Diamond, P. A. (2003). *Taxation, incomplete markets and social security*. MIT Press, Cambridge, MA, The 2000 Munich Lectures edition.

Diamond, P. A. and Spinnewijn, J. (2011). Capital Income Taxes with Heterogeneous Discount Rates. *American Economic Journal: Economic Policy*, 3(November 2011):52–76.

Fleurbaey, M. (1994). On fair compensation. *Theory and Decision*, 36:277–307.

Fleurbaey, M., Leroux, M.-L., and Ponthiere, G. (2014). Compensating the dead. *Journal of Mathematical Economics*, 51:28 – 41.

Fleurbaey, M. and Maniquet, F. (2006). Fair Income Tax. *The Review of Economic Studies*, 73:55–83.

Golosov, M., Troshkin, M., Tsyvinski, A., and Weinzierl, M. (2013). Preference heterogeneity and optimal capital income taxation. *Journal of Public Economics*, 97:160–175.

Lockwood, B. B. and Weinzierl, M. (2015). De Gustibus non est Taxandum: Heterogeneity in preferences and optimal redistribution. *Journal of Public Economics*, 124:74–80.

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimal Income Taxation. *The Review of Economic Studies*, 38(2):175–208.

Roemer, J. E. (1998). *Equality of Opportunity*. Harvard University Press.

Saez, E. (2002). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics*, 83:217–230.

Sandmo, A. (1993). Optimal Redistribution when Tastes differ. *Finanzarchiv*, 50:149–163.

Stern, N. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, 17:181–211.

Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics*, 17:213–240.

Tarkiainen, R. and Tuomala, M. (1999). Optimal Nonlinear Income Taxation with a Two-Dimensional Population; A Computational Approach. *Computational Economics*, 13:1–16.

Tarkiainen, R. and Tuomala, M. (2007). On Optimal Income Taxation with Heterogeneous Work Preferences. *International Journal of Economic Theory*, 3:35–46.

Tenhunen, S. and Tuomala, M. (2010). On optimal lifetime redistribution policy. *Journal of Public Economic Theory*, 12(1):171–198.

Tuomala, M. and Tenhunen, S. (2013). On the design of an optimal non-linear tax/pension system with habit formation. *International Tax and Public Finance*, 20:485–512.

Van de gaer, D. (1993). *Equality of opportunity and investment in human capital*. Phd dissertation, Katholieke Universiteit Leuven.

Appendix A: First-order conditions

To shorten the notation, we denote the partial derivatives as follows:
$\frac{du(c^i)}{dc^i} = u_c^i$, $\frac{dv(x^i)}{dx^i} = v_x^i$ and $\frac{d\psi(1-y^i)}{dy^i} = \psi'^i$

Two types

The Lagrange of the optimization problem is

$$\mathcal{L} = N^L[u(c^L) + \delta^L v(x^L) + \psi(1-y^L)] + \lambda[\sum N^i(n^i y^i - c^i - rx^i) - R] \\ + \mu^{HL}\left[u(c^H) + \delta^H v(x^H) + \psi(1-y^H) - u(c^L) - \delta^H v(x^L) - \psi(1 - \frac{n^L}{n^H}y^L)\right]. \tag{A.1}$$

The first-order conditions with respect to c^i, x^i and y^i, $i = L, H$ are

$$N^L u_c^L - \lambda N^L - \mu^{HL} u_c^L = 0 \tag{A.2}$$
$$N^L \delta^L v_x^L - \lambda r N^L - \mu^{HL} \delta^H v_x^L = 0 \tag{A.3}$$
$$-N^L \psi'^L + \lambda N^L n^L + \mu^{HL} \frac{n^L}{n^H} \psi'^L = 0 \tag{A.4}$$
$$-\lambda N^H + \mu^{HL} u_c^H = 0 \tag{A.5}$$
$$-\lambda r N^H + \mu^{HL} \delta^H v_x^H = 0 \tag{A.6}$$
$$\lambda N^H n^H - \mu^{HL} \psi'^H = 0 \tag{A.7}$$

Three types: Low-productivity workers pooled

Using the information of the binding self-selection constraints provided by numerical solution, the Lagrange function in the case of maximin objective function can be written as

$$
\begin{aligned}
\mathcal{L} = N^1[u(c^L) + \delta^L v(x^L) + \psi(1-y^L)] + \lambda[\textstyle\sum N^i(n^i y^i - c^i - rx^i) - R] \\
+ \mu^{43}[u(c^4) + \delta^H v(x^4) + \psi(1-y^4) - u(c^3) - \delta^H v(x^3) - \psi(1-y^3)] \\
+ \mu^{31}[u(c^3) + \delta^L v(x^3) + \psi(1-y^3) - u(c^1) - \delta^L v(x^1) - \psi(1 - \frac{n^L}{n^H} y^1)] \quad \text{(A.8)}
\end{aligned}
$$

The first-order conditions with respect to c^i, x^i and y^i, $i = 1, 3, 4$ are given by

$$N^1 u_c^1 - \lambda N^1 - \mu^{31} u_c^1 = 0 \tag{A.9}$$
$$N^1 \delta^L v_x^1 - \lambda r N^1 - \mu^{31} \delta^L v_x^{31} = 0 \tag{A.10}$$
$$N^1 \psi' - \lambda N^1 n^L - \mu^{31} \frac{n^L}{n^H} \psi'^{31} = 0 \tag{A.11}$$
$$-\lambda N^3 - \mu^{43} u_c^{43} + \mu^{31} u_c^3 = 0 \tag{A.12}$$
$$-\lambda r N^3 - \mu^{43} \delta^H v_x^{43} + \mu^{31} \delta^L v_x^3 = 0 \tag{A.13}$$
$$-\lambda N^3 n^H - \mu^{43} \psi'^{43} + \mu^{31} \psi'^{31} = 0 \tag{A.14}$$
$$-\lambda N^4 + \mu^{43} u_c^4 = 0 \tag{A.15}$$
$$-\lambda r N^4 + \mu^{43} \delta^H v_x^4 = 0 \tag{A.16}$$
$$-\lambda N^4 n^H + \mu^{43} \psi' = 0 \tag{A.17}$$

In the utilitarian case, the Lagrange function with binding incentive-compatibility constraints can be written as

$$
\begin{aligned}
\mathcal{L} = \textstyle\sum N^i[u(c^i) + \delta^i v(x^i) + \psi(1-y^i)] + \lambda[\textstyle\sum N^i(n^i y^i - c^i - rx^i) - R] \\
+ \mu^{43}[u(c^4) + \delta^H v(x^4) + \psi(1-y^4) - u(c^3) - \delta^H v(x^3) - \psi(1-y^3)] \\
+ \mu^{31}[u(c^3) + \delta^L v(x^3) + \psi(1-y^3) - u(c^1) - \delta^L v(x^1) - \psi(1 - \frac{n^L}{n^H} y^1)].
\end{aligned} \tag{A.18}
$$

The first order condition with respect to c^i, x^i and y^i, $i = 1, 3, 4$ are given by

$$N^1 u_c^1 - \lambda N^1 - \mu^{31} u_c^1 = 0 \tag{A.19}$$
$$N^1 \delta^L v_x^1 - \lambda r N^1 - \mu^{31} \delta^L v_x^{31} = 0 \tag{A.20}$$
$$N^1 \psi'^1 - \lambda N^1 n^L - \mu^{31} \frac{n^L}{n^H} \psi'^{31} = 0 \tag{A.21}$$
$$N^3 u_c^3 - \lambda N^3 - \mu^{43} u_c^{43} + \mu^{31} u_c^3 = 0 \tag{A.22}$$
$$N^3 \delta^L v_x^3 - \lambda r N^3 - \mu^{43} \delta^H v_x^{43} + \mu^{31} \delta^L v_x^3 = 0 \tag{A.23}$$
$$N^3 \psi'^3 - \lambda N^3 n^H - \mu^{43} \psi'^{43} + \mu^{31} \psi'^{31} = 0 \tag{A.24}$$
$$N^4 u_c^4 - \lambda N^4 + \mu^{43} u_c^4 = 0 \tag{A.25}$$
$$N^4 \delta^H v_x^4 - \lambda r N^4 + \mu^{43} \delta^H v_x^4 = 0 \tag{A.26}$$
$$N^4 \psi'^4 - \lambda N^4 n^H + \mu^{43} \psi'^{43} = 0 \tag{A.27}$$

The distortions in this case are

$$\begin{aligned} d^1 &= 0 \\ d^3 &= \frac{\mu^{43}}{N^3 + \mu^{31} - \mu^{43}} \Delta^{HL} \\ d^4 &= 0. \end{aligned} \tag{A.28}$$

Three types: high-productivity workers pooled

Using the information of the binding self-selection constraints provided by numerical solution, the Lagrange function can be written as

$$
\begin{aligned}
\mathcal{L} = N^1[u(c^1) + \delta^L v(x^1) + \psi(1-y^1)] + N^2[u(c^2) + \delta^H v(x^2) + \psi(1-y^2)] \\
+ \lambda[\sum_{i=1}^{n} N^i(n^i y^i - c^i - rx^i) - R] \\
+ \mu^{21}\left[u(c^2) + \delta^H v(x^2) + \psi(1-y^2) - u(c^1) - \delta^H v(x^1) + \psi(1-y^1)\right] \\
+ \mu^{41}\left[u(c^4) + \delta^H v(x^4) + \psi(1-y^4) - u(c^1) - \delta^H v(x^1) + \psi(1-\frac{n^L}{n^H}y^1)\right] \\
+ \mu^{42}\left[u(c^4) + \delta^H v(x^4) + \psi(1-y^4) - u(c^2) - \delta^H v(x^2) + \psi(1-\frac{n^L}{n^H}y^2)\right]
\end{aligned}
\tag{A.29}
$$

The first-order conditions with respect c^i, x^i and y^i, $i = 1, 2, 4$ are given by

$$N^1 u_c^1 - \lambda N^1 - \mu^{21} u_c^1 - \mu^{41} u_c^1 = 0 \tag{A.30}$$

$$N^1 \delta^L v_x^1 - \lambda r N^1 - \mu^{21} \delta^H v_x^1 - \mu^{41} \delta^H v_x^1 = 0 \tag{A.31}$$

$$-N^1 \psi_y^1 + \lambda N^1 n^L + \mu^{21} \psi_y^1 + \mu^{41} \frac{n^L}{n^H} \psi_y^1 = 0 \tag{A.32}$$

$$N^2 u_c^2 - \lambda N^2 + \mu^{21} u_c^2 - \mu^{42} u_c^2 = 0 \tag{A.33}$$

$$N^2 \delta^H v_x^2 - \lambda r N^2 + \mu^{21} \delta^H v_x^2 - \mu^{42} \delta^H v_x^2 = 0 \tag{A.34}$$

$$-N^2 \psi_y^2 + \lambda N^2 n^L - \mu^{21} \psi_y^2 + \mu^{42} \frac{n^L}{n^H} \psi_y^2 = 0 \tag{A.35}$$

$$-\lambda N^4 + \mu^{42} u_c^4 + \mu^{41} u_c^4 = 0 \tag{A.36}$$

$$-\lambda r N^4 + \mu^{42} \delta^H v_x^4 + \mu^{41} \delta^H v_x^4 = 0 \tag{A.37}$$

$$-\lambda N^4 n^H - \mu^{42} \psi_y^4 - \mu^{41} \psi_y^4 = 0 \tag{A.38}$$

Four types

Using the information of the binding self-selection constraints provided by numerical solution, the optimization problem can be written as

$$
\begin{aligned}
\mathcal{L} = N^1[u(c^1) + \delta^L v(x^1) + \psi(1-y^1) + N^2[u(c^2) + \delta^H v(x^2) + \psi(1-y^2) \\
+ \lambda[\sum_{i=1}^{n} N^i(n^i y^i - c^i - rx^i) - R] \\
+ \mu^{31}\left[u(c^3) + \delta^L v(x^3) + \psi(1-y^3) - u(c^1) - \delta^L v(x^1) + \psi(1 - \frac{n^L}{n^H}y^1)\right] \\
+ \mu^{32}\left[u(c^3) + \delta^L v(x^3) + \psi(1-y^3) - u(c^2) - \delta^L v(x^2) + \psi(1 - \frac{n^L}{n^H}y^2)\right] \\
+ \mu^{42}\left[u(c^4) + \delta^H v(x^4) + \psi(1-y^4) - u(c^2) - \delta^H v(x^2) + \psi(1 - \frac{n^L}{n^H}y^2)\right] \\
+ \mu^{43}\left[u(c^4) + \delta^H v(x^4) + \psi(1-y^4) - u(c^3) - \delta^L v(x^3) + \psi(1-y^3)\right]
\end{aligned}
\tag{A.39}
$$

The first-order conditions w.r.t c^i, x^i, and y^i for $i = 1, 2, 3, 4$ are

$$N^1 u_c^1 - \lambda N^1 - \mu^{31} u_c^1 = 0 \tag{A.40}$$

$$N^1 \delta^L v_x^1 - \lambda r N^1 - \mu^{31} \delta^L v_x^1 = 0 \tag{A.41}$$

$$-N^1 \psi_y^1 + \lambda N^1 n^L + \mu^{31} \frac{n^L}{n^H} \psi_y^1 = 0 \tag{A.42}$$

$$N^2 u_c^2 - \lambda N^2 - \mu^{32} u_c^2 - \mu^{42} u_c^2 = 0 \tag{A.43}$$

$$N^2 \delta^H v_x^2 - \lambda r N^2 - \mu^{32} \delta^L v_x^2 - \mu^{42} \delta^H v_x^2 = 0 \tag{A.44}$$

$$-N^2 \psi_y^2 + \lambda N^2 n^L + \mu^{32} \frac{n^L}{n^H} \psi_y^2 + \mu^{42} \frac{n^L}{n^H} \psi_y^{\prime 2} = 0 \tag{A.45}$$

$$-\lambda N^3 + \mu^{31} u_c^3 + \mu^{32} u_c^3 - \mu^{43} u_c^3 = 0 \tag{A.46}$$

$$-\lambda r N^3 + \mu^{31} \delta^L v_x^3 + \mu^{32} \delta^L v_x^3 - \mu^{43} \delta^H v_x^3 = 0 \tag{A.47}$$

$$-\lambda N^3 n^H - \mu^{31} \frac{n^L}{n^H} \psi_y^3 - \mu^{32} \frac{n^L}{n^H} \psi_y^3 + \mu^{43} \psi_y^3 = 0 \tag{A.48}$$

$$-\lambda N^4 + \mu^{42} u_c^4 + \mu^{43} u_c^4 = 0 \tag{A.49}$$

$$-\lambda r N^4 + \mu^{42} \delta^H v_x^4 + \mu^{43} \delta^H v_x^4 = 0 \tag{A.50}$$

$$-\lambda N^4 n^H - \mu^{42} \psi_y^4 - \mu^{41} \psi_y^4 = 0 \tag{A.51}$$

Appendix B: Additional results from the numerical simulations

Table B1 Lagrange multipliers and average tax rates for two-type model, maximin case. Binding constraints in optimum are bolded. For non-binding constraint the value of the constraint $(U^{ij} - U^{i})$ is given in paranthesis.

λ	μ^{LH}	μ^{HL}		Type L	Type H
0.984	0 (-4.53)	0.278	Average tax rate	-40.5	17.0

Table B2 Lagrange multipliers and average tax rates for two-type model in utilitarian case. Binding constraints in optimum are bolded. For non-binding constraint the value of the constraint $(U^{ij} - U^{i})$ is given in paranthesis.

λ	μ^{LH}	μ^{HL}		Type L	Type H
1.87	0 (-2.69)	0.08	Average tax rate	-16.5	8.1

Table B3 Lagrange multipliers and average tax rates for three-type model in maximin case. Pooling of low-ability types. Binding constraints in optimum are bolded. For non-binding constraint the value of the constraint $(U^{ij} - U^{i})$ is given in paranthesis.

λ	μ^{13}	μ^{14}	μ^{31}	μ^{34}	μ^{41}	μ^{43}
1.02	0 (-2.15)	0 (-3.6)	0.33	0 (-0.07)	0 (-0.07)	0.15
	Type 1	Type 3	Type 4			
Average tax rate	-26.4	12.7	14.6			

Table B4 Lagrange multipliers and average tax rates for three-type model in maximin case. Pooling of high-ability types. Binding constraints in optimum are bolded. For non-binding constraint the value of the constraint $(U^{ij} - U^{i})$ is given in paranthesis.

λ	μ^{12}	μ^{14}	μ^{21}	μ^{24}	μ^{41}	μ^{42}
1.03	0 (-0.07)	0 (-3.8)	0.02	0 (-3.3)	0.12	0.18
	Type 1	Type 2	Type 4			
Average tax rate	-32.7	-27.4	15.2			

Table B5 Lagrange multipliers and average tax rates for three-type model in utilitarian case. Pooling of low-ability types. Binding constraints in optimum are bolded. For non-binding constraint the value of the constraint $(U^{ij} - U^{i})$ is given in paranthesis.

λ	μ^{13}	μ^{14}	μ^{31}	μ^{34}	μ^{41}	μ^{43}
1.85	0 (-1.97)	0 (-2.65)	0.11	0 (-0.04)	0(-0.08)	0.04
	Type 1	Type 3	Type 4			
Average tax rate	-17.9	3.4	3.7			

Table B6 Lagrange multipliers and average tax rates for three-type model in utilitarian case. Pooling of high-ability types. Binding constraints in optimum are bolded. For non-binding constraint the value of the constraint $(U^{ij} - U^{i})$ is given in paranthesis.

λ	μ^{12}	μ^{14}	μ^{21}	μ^{24}	μ^{41}	μ^{42}
1.92	0(-0.03)	0 (-2.8)	0.02	0 (-2.65)	0.12	0.18
	Type 1	Type 2	Type 4			
Average tax rate	-19.6	-19.3	4.7			

3 ON OPTIMAL INCOME TAXATION WHEN INHERITED WEALTH DIFFERS

Terhi Ravaska, *University of Tampere & Labour Institute for Economic Research, Helsinki, Finland*
Matti Tuomala, *University of Tampere, Tampere , Finland*

Abstract

In this essay[1] *we study a multidimensional optimal taxation problem when individuals have differences in skills and in initial wealth. In a two-period model with one cohort we derive the optimal distortions for the saving decision in two- to four-types economies. The government aims to redistribute income from the high-income and high-inheritance type towards the low-income and low-inheritance type and to set up a tax system that creates incentives for agents to reveal their true types. Numerical methods are used for solving the binding incentive constraints and optimal consumption-saving-and-work bundles. We also extend the model to include income shifting.*

Our findings support the view that there should be non-linear capital income tax. In the simplest case of two-types, the saving decisions of the low-ability and low-wealth type is taxed at the margin. In the 3- to 4-type settings high initial wealth types are subsidized at the margin. The subsidy relax the self-selection constraint which prevent the high-wealth types mimicking to be low-wealth types. For the type of low-wealth and high-productivity the marginal distortion on the saving decision depends upon the degree of correlation between ability and initial wealth and the chosen social welfare function.

[1]We thank Albert Jan Hummel for the comments for the earlier version of this paper. We also thank the seminar participants in the IIPF conference 2016 in Lake Tahoe and at the Tampere Allecon meeting in 2016.

Keywords: Optimal taxation, lifetime redistribution, multidimensional tax problem, heterogeneity in inherited wealth, income shifting
JEL classification: H21, D61, D71

3.1 Introduction

In many developed countries one highly significant phenomenon in recent years has been the ending of the downward trend in wealth concentration. Piketty and Zucman (2014) have estimated wealth-to-income ratios for eight advanced economies and their estimates reveal some striking trends. Wealth to income in these nations climbed from a range of 200 to 300 percent in 1970 to a range of 400 to 600 percent in 2010. The wealth differences for any given cohort will reflect income differences if individuals save for life-cycle smoothing purposes and everyone has the same preferences. However, this is not the only way in which people receive capital since some people inherit it. Hence, capital income inequality stems from differences in wealth due to past saving behaviour, inheritances received, and in rates of return that have varied dramatically over time and across assets.

Mirrlees (1971) states that *"In an optimum system, one would no doubt wish to relate tax payments to the whole life pattern of income, and to initial wealth".* In practice, taxation is not based on life-cycle but on annual incomes and initial wealth differences are only partly accounted in inheritance taxation. Especially the initial wealth differences have received only little attention as a source of heterogeneity in the optimal income taxation literature. Most of the optimal income taxation literature has focused solely on differences in productivities, and only recently has heterogeneity in other dimensions[2] been incorporated into the models.

In the optimal inheritance taxation literature the center of interest is to determine how to tax the bequests left behind usually by parents to their children. This is a one-off occasion from the perspective of taxation. There is an ample set of models studying this issue, taking into account different as-

[2]For example time preference differences have been incorporated into optimal taxation models by Tenhunen and Tuomala (2010) and Diamond and Spinnewijn (2011), myopia by Cremer et al. (2009) and differences on disutility of labour by Boadway et al. (2002).

sumptions about preferences for saving and bequest (see Kopczuk (2013) and Piketty and Saez (2013) and the references therein). In these studies the optimal tax rates vary considerably from zero to extreme high. However, in this paper we are not interested in taxation of this kind of one-off event but instead focus on the question how initial wealth affect the structure of income taxation under different redistributive preferences. Especially we study whether savings should be taxed or not. This approach is also motivated by the empirical fact that inheritance tax is one of the least popular forms of taxes and many countries have abolished it all together with other net wealth taxes (Drometer and Frank, 2018).

In our paper, people differ in terms of productivity and in initial wealth which are both unobservable to the policy-maker[3]. Initial endowments put individuals in a different starting point in their life already before the productivity-type is revealed. In a world without initial wealth differences, redistribution would occur only between productivity types but with initial endowments the direction of redistribution between types is ambiguous. To learn more about the direction of redistribution we solve numerical examples.

We study optimal taxation in a static setting in a sense that we take the endowment or initial wealth as exogenous. This means that we study only one cohort who are entitled to a bequest from the previous cohort but leave no bequest. The focus is put on the optimal distortion for savings and hence for the question whether or not to tax capital income but we also comment on the optimal levels of the labour supply distortions in the numerical simulations. We contribute to the earlier literature by studying non-linear savings tax and extend the model to include also income shifting.

Our analytical results (assuming the direction of binding self-selection constraints) and numerical solution reveal that the saving decision is distorted at the margin if there are differences in initial wealth between individuals. The common pattern in most specifications is that there is a tax at the margin for the individuals with low initial wealth and a subsidy for the individuals with high initial wealth. In the numerical simulations we also consider the role of wealth inequality, correlation between ability and wealth and different social

[3]In reality some of the bequests are observable to the government but there are also transfers that are either unobservable or unidentifiable.

objectives. Wider wealth inequality requires less distortions for the saving decision. The correlation between the unobserved factors affect the optimal distortion in a non-monotonic way and the social objectives matter for which type is distorted at the margin.

The rest of the paper is organized as follows. In section 3.2 we discuss the earlier literature. In section 3.3 we introduce the benchmark model. In section 3.4 we extend the type-space to three and four types and discuss the numerical simulations. In section 3.5 we extend the model to include income shifting. Section 3.6 concludes.

3.2 Earlier literature

For a recent review of the literature of capital taxation see Bastani and Waldenström (2018). Wealth or inheritance differences between agents have mainly been discussed in the literature of optimal inheritance taxation. This literature provides inconclusive results whether and how inheritances should be taxed. It seems that the motive of bequest is central and whether bequest is voluntary or accidental.

The Atkinson-Stiglitz result (Atkinson and Stiglitz, 1976) says that by assuming a mild separability between consumption and labour supply, the non-linear labour income tax does not need to be supplemented with other taxes, like savings tax. There are very few papers on the role of capital income taxation in the Atkinson-Stiglitz type of economy with differences in unobserved inherited wealth. Such economies are studied in Boadway et al. (2000) and Cremer et al. (2003). Boadway et al. (2000) study overlapping generations model where savings tax is linear and labour income tax non-linear. They show that with accidental bequests, tax on the interest income can indirectly tax the unobserved inherited wealth and is thus desirable. Cremer et al. (2003) derive the same result with another bequest motive, 'joy of giving'. They conclude that capital income tax can be a desirable additional instrument to bring more information about the unobserved inheritances. With positive correlation between inherited wealth and productivity, marginal tax rate on capital income is likely a positive one.

Many countries have abolished wealth and inheritance taxes (Drometer and Frank, 2018). This motivates us to study initial endowment differences also in a framework where the focus is on one cohort. The same kind of economy is studied for example in Cremer et al. (2001) and Christiansen and Tuomala (2008). These papers study whether linear capital income tax should be added to the optimal tax mix. Our study differs from these two studies by studying non-linear capital income tax.

Non-linear capital income taxation is relatively understudied. Capital income taxation is often studied in the framework of commodity taxation and for many commodities the non-linear taxation is impractical for the tax arbitrage reasons. However, we argue that non-linear capital income taxation is possible especially in developed countries where income is often reported by third-parties. Several countries have tax policies that effectively vary the capital income tax rate based on the total income (USA) or annual capital income (Finland). This makes our case for studying this tax instrument. For other studies on non-linear capital income taxation, see for example Golosov et al. (2013).

3.3 Benchmark model

We assume that individuals differ in their productivity and in their initial wealth level. First we consider a simple two-type economy where productivity and initial wealth are perfectly correlated. Table 3.1 shows evidence for Finland that wealth and capital income indeed are strongly correlated and the correlation has got stronger over time making this a reasonable assumption for the benchmark model. In later parts this assumption of perfect correlation is relaxed.

Table 3.1 Correlation between income items and net wealth. Data sources:Wealth Study Statistic Finland. Source Riihelä et al. (2007), updated.

	Correlations in 1987			*Correlations in 2013*		
	Net wealth	Labour income	Capital income	Net wealth	Labour income	Capital income
Net wealth	1.000			1.000		
Labour income	0.219	1.000		0.106	1.000	
Capital income	0.360	0.145	1.000	0.743	0.060	1.000

In our model the initial wealth is exogenous lump-sum, which is received in the first period before labour supply decisions are made. Types are exogenous and the distribution is known by the tax planner. This exogeneity assumption is done to simplify the exposition and notation without loss of generality. There is asymmetric information in a sense that the tax authority can observe neither the endowments, productivities nor labour supply and this rules out first best taxation. Tax authorities can observe the savings from the first period to the second period and the earned income in the second period. These assumptions are made in order to study a fully non-linear tax system.

The government wants to create a lifetime tax system which redistributes income between the individuals in the same cohort and we assume it can commit to the chosen tax-and-transfer system. The government can now use both labour income tax and capital income tax. In this setting we study whether capital income tax is needed in the optimal tax mix.

3.3.1 Individuals

Each individual has a skill level, n^i, reflecting his wage level and an initial wealth level, e^i reflecting the consumption potential in the first period. Low-skilled and/or low-endowed are denoted with superscript L and high-skilled and/or high-endowed are denoted with superscript H. The benchmark assumption of the positive correlation implies that we have two types of (e^H, n^H) and (e^L, n^L). The proportion of each type i in the population is $N^i \geq 0$ and $\sum N^i = 1$.

We consider a two-period model. In order to say something about the Atkinson-Stiglitz result, the life-time utility of an agent i is separable and additive in the following way:

$$U^i = u(c^i) + \delta v(x^i) + \psi(1 - y^i), \tag{3.1}$$

where c and x denote consumption in the first and the second period, respectively, and y is the labour supply during the second period. δ denotes the discount factor. It is assumed that U is a strictly concave, continuously differentiable and utility is strictly increasing in c and x, and strictly decreasing in

y, and (partial derivatives) u', v', ψ', $\psi'' > 0$ and u'', $v'' < 0$. We also assume that all goods are normal and technologies are linear, that is one extra unit of labour produces one unit of commodity good. Markets are assumed to be competitive.

In the first period individuals divide the initial wealth between consumption, c, and saving, s. Each unit of savings yields an additional $1+\theta$ units of consumption in the second period. We simplify the analysis by considering a small open economy facing world capital markets which implies that the return to savings, θ, is fixed. Consumption in the first period is $c^i = e^i - s^i$ and in the second period $x^i = (1+\theta)s^i + B^i$, where after-tax labour income is denoted by $B^i = n^i y^i - T(n^i y^i)$. The inter-temporal budget constraint then is $c^i + rx^i = e^i + rB^i$, where $r = \frac{1}{1+\theta}$.

Individual's problem is to maximize inter-temporal utility with budget constraint:

$$\begin{gathered} \max \quad u(c^i) + v(x^i) + \psi(1-y^i) \\ \text{s.t.} \\ c^i + rx^i = e^i + rB^i. \end{gathered} \tag{3.2}$$

Without distortive taxation the well-known Euler equation emerges from the first order conditions:

$$\frac{u'}{v'} = \frac{\delta}{r}. \tag{3.3}$$

We can also solve the familiar formula for marginal labour income tax:

$$\frac{\psi'}{\delta n v'} = 1 - T'. \tag{3.4}$$

3.3.2 Government

The government has aversion towards inequality so it creates a tax-and-transfer system which redistributes income from better-off towards worse-off. In this setting individuals state their initial wealth and productivity type to the government and the government offers them a bundle of gross and net incomes. The self-selection or incentive constraints make sure that the individual weakly prefers the bundle aimed at him over all the other bundles.

In the benchmark case we assume that there is a perfect correlation between productivity and endowment received. Here the single-crossing property holds and the only binding self-selection constraint is from the direction of high-type towards low-type. In preceding sections, when the perfect correlation assumption is relaxed, numerical methods are used for determining the binding self-selection constraints. The incentive constraints which prevent the mimicking behaviour are written as:

$$u(c^H) + \delta v(x^H) + \psi(1 - y^H) \geq \widehat{u}(c^L) + \delta v(x^L) + \psi(1 - \frac{n^L}{n^H} y^L) \equiv U^{ij}, \quad (3.5)$$

where hat denotes mimicking behaviour and specifically $\widehat{u}(c^L) = u(e^H - s^L)$. That is, in the case of mimicking the mimicker would enjoy higher first period consumption than truly low-ability type.

In the case of utilitarian social objective function government's problem is to

$$\begin{aligned} &\underbrace{\max}_{c^i, x^i, y^i} G(U) = \sum_{i=1}^{N} N^i (u(c^i) + \delta v(x^i) + \psi(1 - y^i)) \\ &\qquad \text{subject to} \\ &\qquad \sum_{i=1}^{N} N^i (n^i y^i + (1 + \theta)(e^i - c^i) - x^i) \geq G \\ &\qquad U^H \geq U^{HL}, \end{aligned} \quad (3.6)$$

where the first constraint is the resource constraint and second the aforementioned incentive constraint. An alternative social objective function is maximin, where government maximizes only the welfare of the low-ability type (defined as the individuals with low inheritance and productivity).

3.3.3 Analytical results

We can now derive the first order conditions by forming a Lagrangian of the government's problem. The multiplier for the resource constraint is λ and for the self-selection constraint μ^{HL}. The Lagrangian expression is:

$$
\begin{aligned}
\mathcal{L} &= \sum_{i=1}^{N} N^i[u(c^i) + \delta v(x^i) + \psi(1-y^i)] \\
&+ \lambda[\sum_{i=1}^{N} N^i(n^i y^i + (1+\theta)(e^i - c^i) - x^i) - G] \\
&+ \mu^{HL}[u(c^H) + \delta v(x^H) + \psi(1-y^H) - \widehat{u}(c^L) - \delta v(x^L) - \psi(1 - \frac{n^L}{n^H} y^L)].
\end{aligned} \tag{3.7}
$$

The first order conditions with respect to c^i, x^i and y^i are given by

$$\frac{\partial \mathcal{L}}{\partial c^L} = N^L u' - \lambda N^L(1+\theta) - \mu^{HL}\widehat{u}' = 0 \tag{3.8}$$

$$\frac{\partial \mathcal{L}}{\partial x^L} = N^L \delta v' - \lambda N^L - \mu^{HL}\delta v' = 0 \tag{3.9}$$

$$\frac{\partial \mathcal{L}}{\partial y^L} = -N^L \psi' + \lambda N^L n^L + \mu^{HL}\psi' \frac{n^L}{n^H} = 0 \tag{3.10}$$

$$\frac{\partial \mathcal{L}}{\partial c^H} = N^H u' - \lambda N^H(1+\theta) - \mu^{HL} u' = 0 \tag{3.11}$$

$$\frac{\partial \mathcal{L}}{\partial x^H} = N^H \delta v' - \lambda N^H - \mu^{HL}\delta v' = 0 \tag{3.12}$$

$$\frac{\partial \mathcal{L}}{\partial y^H} = -N^H \psi' + \lambda N^H n^H - \mu^{HL}\psi' = 0. \tag{3.13}$$

In the case of maximin social objective function the last three equations become:

$$\frac{\partial \mathcal{L}}{\partial c^H} = -\lambda N^H(1+\theta) - \mu^{HL} u' = 0 \tag{3.14}$$

$$\frac{\partial \mathcal{L}}{\partial x^H} = -\lambda N^H - \mu^{HL}\delta v' = 0 \tag{3.15}$$

$$\frac{\partial \mathcal{L}}{\partial y^H} = \lambda N^H n^H - \mu^{HL}\psi' = 0. \tag{3.16}$$

It should be noted that the consumption of the mimicker is $\Delta e + c_L$, where $\Delta e = e^H - e^L$. Define then the inter-temporal marginal substitution as $\frac{u'_i}{v'_i}$ and rearrange to get for the high-ability type

$$\frac{u'}{v'} = \frac{\delta}{r}, \tag{3.17}$$

that is, the saving decision is not distorted at the margin even when there are differences in initial wealth. However, for the low-ability type, the saving decision is not distorted only in the case where there is no differences in initial endowments (and so $u'_L = \widehat{u}'$) or when incentive-constraint μ^{HL} is not binding. We can see this by rearranging the first order conditions as before:

$$\frac{N^L u' - \mu^{HL}\widehat{u}'}{v'(N^L - \mu^{HL})} = \frac{\delta}{r}. \tag{3.18}$$

Rearranging equation 3.18 to $\frac{u'}{v'} = \frac{\delta}{r}(1 - d_L)$ and solving for d_L, we get the marginal distortion on the low-types saving decision (assuming the only binding constraint is from high-type towards the low-type):

$$d_L = \frac{\mu^{HL}}{N^L}(1 - \frac{r}{\delta}\frac{\widehat{u}'}{v'}). \tag{3.19}$$

While equation 3.19 does not instantly reveal the sign of the distortion, we know that mimicker's marginal utility of consuming during the first period is smaller than for the true low ability types. In order to prevent the mimicker for consuming too much during the first period, we need to distort the low-ability types saving decision which would make mimicking behaviour less beneficial. That means that the distortion is positive in the current case.

Solving for the optimal marginal labour tax reveals the typical result that the high-ability type's labour decision is not distorted at the margin. As the labour supply decision is independent of the endowment we know that the common case of positive marginal labour income tax takes place for the low-ability type. This distortion equals

$$T'_L = \frac{\mu_{HL}(1 - \frac{n^L}{n^H})}{N^1 - \mu^{HL}}. \tag{3.20}$$

We can state the following proposition:

Proposition 1

In the two-type economy
(i) Both in utilitarian and maximin cases if $e^L = e^H$ and if preferences of individuals are additively separable, then there is no taxation of capital income at the optimum.

Low-ability type faces a positive marginal labour income tax.
(ii) If $e^L < e^H$ and if preferences of individuals are additively separable and the single-crossing condition holds, then there is a case for taxing capital income at the optimum. The low-ability type faces a positive capital income tax at the margin for both utilitarian or maximin social objectives. Low-ability type faces also a positive marginal labour income tax.

3.4 Extenstion to type-space

Relaxing the assumption of perfect correlation between productivity and initial wealth leads to four-type economy as in the table 3.2. With zero correlation, there are equal amount of individuals in each type and for imperfect correlation cases the sizes of the types vary. Including more types leads to larger set of potentially binding self-selection constraints. In the next subsection we study 3-type case by assuming some of the binding constraint while the subsequent subsection utilizes numerical methods for the 4-type model. The government's objective is either utilitarian or maximin with respect of productivity.

Table 3.2 Definition of types by initial wealth and productivity

	e^L	e^H
n^L	1	2
n^H	3	4

The Lagrange function in general form is:

$$\begin{aligned}\mathcal{L} = &\sum_{i=1}^{N} N^i(u(c^i) + \delta v(x^i) + \psi(1-y^i)) + \lambda[\sum_{i=1}^{N} N^i(n^i y^i + (1+\theta)(e^i - c^i) - x^i) - G] \\ &+ \sum_{i,j=1,i\neq j}^{N} \mu^{ij}[u(c^i) + \delta v(x^i) + \psi(1-y^i) - \widehat{u}(c^j)) - \delta v(x^j) - \psi(1 - \frac{n^i}{n^j} y^i)] \\ &+ \sum_{i,j=1,j\neq i}^{N} \mu^{ji}[u(c^j) + \delta v(x^j) + \psi(1-y^j) - \widehat{u}(c^i) - \delta v(x^i) - \psi(1 - \frac{n^j}{n^i} y^j)]\end{aligned} \tag{3.21}$$

The first order conditions with respect to c^i, x^i and y^i are

$$\frac{\partial \mathcal{L}}{\partial c_i} = N^i u' - \lambda(1+\theta)N^i + \sum_{i,j=1,i\neq j}^{N} \mu^{ij} u' - \sum_{i,j=1,i\neq j}^{N} \mu^{ji} \widehat{u}' = 0 \tag{3.22}$$

$$\frac{\partial \mathcal{L}}{\partial x_i} = N^i \delta v' - \lambda N^i + \sum_{i,j=1,i\neq j}^{N} \mu^{ij} \delta v' - \sum_{i,j=1,i\neq j}^{N} \mu^{ji} \delta v' = 0 \tag{3.23}$$

$$\frac{\partial \mathcal{L}}{\partial y_i} = -N^i \psi' + \lambda N^i n^i - \sum_{i,j=1,i\neq j}^{N} \mu^{ij} \psi' + \sum_{i,j=1,i\neq j}^{N} \mu^{ji} \psi' = 0. \tag{3.24}$$

For the maximin social objectives the first term drops out for types 3 and 4. In this setting the distortions depend on which self-selection constraints are binding. Note particularly that the single-crossing property does not necessarily hold meaning that the self-selection constraint can bind towards both directions simultaneously. Whenever there is at least one binding self-selection constraint between types with different initial wealth, there is a distortion for the saving decision compared to the first best. Formally we see this from the condition

$$\frac{(N^i + \sum_{i,j=1,i\neq j}^{N} \mu^{ij})u' - \sum_{i,j=1,i\neq j}^{N} \mu^{ji} \widehat{u}'}{(N^i + \sum_{i,j=1,i\neq j}^{N} \mu^{ij} - \sum_{i,j=1,i\neq j}^{N} \mu^{ji})v'} = \frac{\delta}{r} \tag{3.25}$$

for all i in the case of utilitarian social objectives and for the types 1 and 2 in the case of maximin social objectives. For types 3 and 4 under maximin the condition is:

$$\frac{\sum_{i,j=1,i\neq j}^{N} \mu^{ij} u' - \sum_{i,j=1,i\neq j}^{N} \mu^{ji} \widehat{u}'}{(\sum_{i,j=1,i\neq j}^{N} \mu^{ij} - \sum_{i,j=1,i\neq j}^{N} \mu^{ji})v'} = \frac{\delta}{r}. \tag{3.26}$$

3.4.1 3-type model

In the four-type model we cannot state which of the self-selection constraints are binding. It is especially difficult to prove whether the constraint between types 2 and 3 is binding upwards or downwards, or both ways. To get further insight from the analytical results we study a 3-type case where economy consists low-ability and low-wealth individuals (type 1) and high-ability individuals with varying initial wealth levels (types 3 and 4). We assume that the binding self-selection constraint are from types with higher resources towards types with lower resources meaning $\mu^{41}, \mu^{43}, \mu^{31} > 0$.

Now we can solve the distortions for savings:

$$d_1 = \frac{\mu^{41} + \mu^{31}}{N^1}(1 - \frac{r}{\delta}\frac{\widehat{u}'}{v'}) > 0 \tag{3.27}$$

$$d_3 = 0 \tag{3.28}$$

$$d_4 = 0. \tag{3.29}$$

Distortions are the same under maximin social objectives[4]. That is with wealth heterogeneity the marginal tax on capital income is positive for the low-ability type in order to relax the otherwise binding self-selection constraints.

The optimal marginal labour income tax is non-zero for low-ability type and 0 otherwise, indicating that there is no distortion at the top. Specifically the marginal labour income tax for type 1 is:

$$T_1' = \frac{(\mu^{41} + \mu^{31})(1 - \frac{n^L}{n^H})}{N^1 - (\mu^{41} + \mu^{31})\frac{n^L}{n^H}} \tag{3.30}$$

We can state the following proposition:

Proposition 2

In the three-type economy, where low-ability types have low initial wealth and the wealth level varies within the high-ability type, only the saving decision of the low-ability type is distorted at the margin both in the utilitarian and the maximin cases. The result hinges on the assumption that self-selection constraints are binding along decreasing resources (combining productivity and initial wealth). The "no distortion at the top" result holds.

[4]Note that the levels of marginal distortions vary between the social objectives because the optimal bundles of consumption and leisure differ.

3.4.2 Numerical illustration

For the full four-type model we solve some numerical examples. We choose a Cobb-Douglas utility function and parameter values following Cremer et al. (2001). The utility function is separable and represented in the following way: $U^i = logc^i + \delta logx^i + log(1 - y^i)$. The parameters for the baseline case are as in the table 3.3.

Table 3.3 Parameters for the numerical simulations, baseline

Fraction of individuals in each group	$N^i = 0,25$ for $i = 1,2,3,4$
Endowments	$e^L = 2, e^H = 10$
Discount rate	$\delta = 0.9$ and $r = 0.95$
Productivities	$n^L = 6, n^H = 9$

Table 3.4 shows the optimization result in a baseline case. The upper panel is for the utilitarian social objective function and lower panel for the maximin case. For the benchmark we report the distribution of savings, labour income, lifetime consumption and the distortions for saving and labour income. The seventh column also presents the life-time utility. The binding self-selection constraints are also listed. In the last column we also show a simple inequality measures which are calculated by taking the absolute difference between the extreme types (types 1 and 4).

For the baseline model in the utilitarian case the only slack self-selection constraints are μ^{12}, μ^{13} and μ^{23} and for the maximin case μ^{12}, μ^{13}, μ^{23}, μ^{34} and μ^{43}. This demonstrates that the intuition from the one-dimensional case does not hold in a multidimensional screening problem.

For example constraint μ^{34} binds even though type 4 has a higher disposable income potential in the first best case without distortive taxation. Also it is interesting that μ^{41} binds as it links the extreme types and indicates that type 1 is taxed at the margin partly to prevent the high-ability high-wealth type from pretending to be both low-ability and low-wealth. Difference in the after-tax incomes between these two types can be reduced only until the self-selection constraint becomes binding.

Turning to the optimal distortions, the simulation exercise suggests that the saving decision is distorted with a positive tax for the low-wealth types at the margin. Types with high-wealth types are instead subsidized. The tax on type 1 relaxes all the self-selection constraints towards the type 1. The subsidy for high wealth types helps to mitigate the binding self-selection constraint as now the potential mimickers do not want to meet the high-wealth types' savings level.

The marginal distortion for the labour supply are somewhat surprising, however, not uncommon in the multidimensional tax problems (Cremer et al., 2001). We observe that the negative marginal tax rates occur to the high-ability and high-wealth type. Comparing the social objectives tells that the saving distortion for type 1 and 4 are substantially affected by changing the redistributive preferences to maximin.

Table 3.4 Baseline model simulation

Utilitarian	s	ny	$c+x$	d	T'	lifetime U	
Type 1	0,55	2,06	2,24	29,0	35,3	-0,25	
Type 2	1,90	1,07	9,48	-17,2	32,3	2,18	difference in lifetime consumption: 8,72
Type 3	0,28	6,0	3,96	0,5	0	0,17	difference in lifetime utility: 2,65
Type 4	1,73	4,05	10,96	-24,0	-34,3	2,40	
Binding SS	14, 21, 24, 31, 32, 34, 41, 42, 43						
Maximin	s	ny	$c+x$	d	T'	lifetime U	
Type 1	0,7	2,22	1,86	45,3	47,4	-0,71	
Type 2	1,66	1,74	9,28	-22,5	57,2	1,71	difference in lifetime consumption: 8,7
Type 3	0,53	5,94	3,19	0	0	-0,21	difference in lifetime utility: 2,67
Type 4	1,52	5,04	10,56	-12,8	-12,5	1,96	
Binding SS	14, 21, 24, 31, 32, 41, 42						

Table presents in the baseline model the optimal savings s, labour income ny, lifetime consuption $c+x$, saving distortion d, marginal labour income tax T' and lifetime utility U. Differences in lifetime consumption and utility are calculated as absolute difference between the extreme types, i.e. 1 and 4. The binding self-selection constraints listed at the bottom.

The impact of wealth inequality

Next we will explore the role of wealth inequality. The overall wealth in the economy is kept constant but the distribution is changed. Other parameters are as in baseline case. The results on the variables we are the most interested in are shown in table 3.5.

It appears that the saving distortion increases for type 1 and 3 as the wealth inequality is reducing. At the same time the subsidy provided for the high-wealth types increases. In the utilitarian case the marginal labour income tax for the type 1 is minimally affected with decreasing wealth inequality. For high-initial-wealth types the optimal labour supply distortions are non-monotone.

Generally, as the wealth differences are narrower, the marginal distortions get larger.

Table 3.5 Impact of wealth inequality

Utilitarian									
e^L	e^H	d^1	d^2	d^3	d^4	T^1	T^2	T^3	T^4
2	10	29,0	-17,2	0,5	-24,0	35,3	32,3	0	-34,3
3	9	41,9	-27,3	2,5	-28,3	45,1	33,2	-0,02	-46,2
5	7	59,8	-73,6	16,6	-51,0	52,4	18,0	-24,4	-34,4
6	6	0	0	0	0	23,15	23,15	0	0
Maximin									
2	10	45,3	-22,5	0	-12,8	47,3	57,3	0	-12,5
3	9	65,2	-45,2	0	-25,3	70,4	45,2	0	-9,7
5	7	68,8	-67,3	14,3	-48,9	63,3	37,3	-7,9	-9,8
6	6	0	0	0	0	43,3	43,3	0	0

Correlation between ability and wealth

In table 3.6 we explore how the correlation between ability and wealth affects the optimal distortions keeping the wealth inequality and other parameters in the baseline specification. The above analysis have assumed zero correlation. Correlation 1 corresponds to the case discussed in section 3.3 and verifies that the only binding constraint is towards the low-ability and low-wealth type. In the two-type case the government can recognize the true type more easily than in a case where the is less than perfect correlation[5]. For this reason the distortion for the type 1 is smaller than with the zero or positive correlation cases. Compared to the linear tax case studied in Cremer et al. (2001), there are remarkable differences in the findings. While with linear savings tax and

[5]The perfect negative correlation is not shown because with this specification neither of the self-selection constraint is binding and so one cannot solve for the optimal distortion.

unobservable savings increasing the correlation indicates a higher tax, with the non-linear case there is no clear pattern but only that the distortions are non-monotone.

Table 3.6 Impact of correlation between ability and wealth

				utilitarian				
Correlation	d^1	d^2	d^3	d^4	T^1	T^2	T^3	T^4
1	22,9	0	0	0	11,7	0	0	0
positive	37,7	-20,2	0	-22,3	52,9	23,7	0	-32,4
0	29,0	-17,2	0,5	-24,0	38,3	32,3	0	- 35,0
negative	12.0	-11.0	2.1	-29.4	10.5	39.0	-1.45	-36.4
				maximin				
1	52,6	0	0	0	35.9		0	0
positive	63,1	-25,3	6,8	-9,8	62,2	54,7	-3,7	-6,7
0	45,3	-22,5	0	-12,8	51,5	57,2	0	-7,1
negative	0	-14,3	0	-18,5	38,4	55,2	0	-7,2

3.5 Including income shifting

Lastly, we extend our model to include income shifting. The same question with similar model has been studied in Christiansen and Tuomala (2008) but with a linear capital income tax. They study income shifting in a two-type and two period model when there is a perfect correlation between productivity and initial wealth. They solve the optimal tax rate analytically. For now, we will also consider a two-type model with perfect correlation and leave the numerical results and other cases of correlations for future. We are interested in whether the results from Christiansen and Tuomala (2008) extend to an economy where government can also tax savings with a non-linear tax schedule.

Earlier we assumed that government can observe savings and income from labour. However, now we shift our focus on more realistic environment where savings and labour income are reported to the government. So, from the point of view of the government, the savings and income are only partially observed. It is costly for government to monitor if incomes are reported truthfully. This creates an incentive to shift part of the labour income to capital income if there are differences in the tax rates.

To model income shifting, we assume that individuals shift an amount of Δ of labour income to capital income tax base at a cost of $k(\Delta)$. We restrict only to the case where $k(\Delta) > 0$ and so rule out the income shifting from capital income to labour income. The cost is increasing and convex in Δ, i.e $k'(\Delta) > 0$, $k''(\Delta) > 0$. Individual report labour income z^R while the true labour income equals to $z = z^R + \Delta$. Individual's actual labour supply is $y = \frac{z^R+\Delta}{n}$. Individuals shift capital income as long as the marginal saving in tax is larger than the cost of transforming one unit of labour income to capital.

Each type chooses how much to report as savings, how much to save from the first period and the labour supply to maximize his utility. To characterize Pareto efficient second best taxes, we set the government problem as to maximize

$$\begin{aligned} &\underbrace{\max}_{c^i,x^i,y^i} G(U) = \sum_{i=1}^{N} N^i(u(c^i) + \delta v(x^i) + \psi(1-y^i)) \\ &\qquad \text{subject to} \\ &\qquad \sum_{i=1}^{N} N^i(n^i y^i + (1+\theta)(e^i - c^i) - x^i) \geq G \\ &\qquad U^H \geq U^{HL}, \end{aligned} \tag{3.31}$$

where self-selection constraint are (assuming only μ^{HL} binds)

$$u(c^H) + \delta v(x^H) + \psi(1-y^H) \geq \widehat{u} + \delta\widehat{v} + \psi(1 - \frac{n^L}{n^H} y^L) \equiv U^{ij}. \tag{3.32}$$

As before $\widehat{u}$ denotes the mimickers first period utility at $u(c^L + e^H - e^L)$. The new type of mimicking in this context is that the high-ability type may want to mimic the low-ability type and in reality have larger second period consumption due to the income shifting. In this case the second period utility equals to $\widehat{v}(x^L + (\Delta^H - k(\Delta^H) - (\Delta^L - k(\Delta^L)))$. Both types of mimicking behaviour needs to be accounted.

Solving the first order conditions and rearranging the terms as before for the saving distortion we notice that the high-ability and high-wealth type faces zero distortion as before but for the low-ability and low-wealth type the distortion is:

$$\frac{N^L u' - \mu^{HL}\widehat{u}'}{N^L v' - \mu^{HL}\widehat{v}'} = \frac{\delta}{r}. \tag{3.33}$$

If there is no wealth differences, the nominator becomes $(N^L - \mu^{HL})u'$. The denominator takes into account the income shifting. Because $\widehat{v}' < v'$ (assuming that high-ability individual shifts more income), the denominator is larger and the marginal rate of substitution between the two periods is smaller than without income shifting. This indicates that there is a positive marginal tax for the low-ability -type. If there are differences in the initial wealth, the sign of the distortion is ambiguous since nominator and denominator is working in opposite directions. This requires numerical simulations and are left for the future versions of the model.

3.6 Conclusions

Rising inequality is often (too often) discussed in differences in labour income. The standard optimal income tax analysis is also based on differences in earnings capacity. The wage distribution is certainly important but people differ also in wealth they have. Increasing wealth inequality has motivated us to study the non-linear taxation of labour and capital income in a two period model where agents differ in their earnings capacity and in their initial wealth levels.

Given the Atkinson-Stiglitz result capital income tax might be unnecessary in a setting where non-linear labour income tax is available. We have shown that there may, however, be a role for taxing savings. Non-linear marginal savings tax schedule alleviates the informational constraints. Our finding indicates that low-wealth types' savings are taxed at margin while the high-wealth type's saving decision is distorted upwards. This reveals that there is a case for non-linear capital income tax in the optimal tax instrument mix.

Because multidimensional problems can be difficult to grasp without numerical simulation, in particular because the analytical results turns out to be ambiguous, several model economies are set up for the numerical examples in this paper. We have studied the effect of wealth inequality on the optimal savings and labour supply distortion. We have also characterized the effect of correlation between ability and wealth. Lastly, we have studied the role of income shifting in an analytical framework.

We motivated the study with the observation that wealth and inheritance taxes have been going out of favour in many developed country. While we have shown that the capital income tax belongs to the optimal tax mix, the extent of redistribution changes little between the maximin and the utilitarian social objectives. This might suggest that initial wealth might be easier to target with wealth taxation.

Bibliography

Atkinson, A. B. and Stiglitz, J. E. (1976). The Design of Tax Structure: Direct versus Indirect Taxation. *Journal of Public Economics*, 6:55–75.

Bastani, S. and Waldenström, D. (2018). How Should Capital Be Taxed? Mimeo.

Boadway, R., Marchand, M., and Pestieau, P. (2000). Redistribution with Unobservable Bequests: A Case for Taxing Capital Income. *Scandinavian Journal of Economics*, 102(2):253–267.

Boadway, R., Marchand, M., Pestieau, P., and Racionero, M. D. M. (2002). Optimal Redistribution with Heterogeneous Preferences for Leisure. *Journal of Public Economic Theory*, 4(4):475–498.

Christiansen, V. and Tuomala, M. (2008). On Taxing Capital Income with Income Shifting. *International Tax and Public Finance*, 15:527–545.

Cremer, H., De Donder, P., Maldonado, D., and Pestieau, P. (2009). Forced Saving, Redistribution, and Nonlinear Social Security Schemes. *Southern Economic Journal*, 76(1):86–98.

Cremer, H., Pestieau, P., and Rochet, J. C. (2001). Direct versus indirect taxation: The design of the tax structure revisited. *International Economic Review*, 42(3):19.

Cremer, H., Pestieau, P., and Rochet, J. C. (2003). Capital income taxation when inherited wealth is not observable. *Journal of Public Economics*, 87(11):2475–2490.

Diamond, P. A. and Spinnewijn, J. (2011). Capital Income Taxes with Heterogeneous Discount Rates. *American Economic Journal: Economic Policy*, 3(November 2011):52–76.

Drometer, M. and Frank, M. (2018). Wealth and Inheritance Taxation: An Overview and Country Comparison. *ifo DICE Report*, 16(2):45–54.

Golosov, M., Troshkin, M., Tsyvinski, A., and Weinzierl, M. (2013). Preference heterogeneity and optimal capital income taxation. *Journal of Public Economics*, 97:160–175.

Kopczuk, W. (2013). Taxation of Intergenerational Transfers and Wealth. In Auerbach, A., Chetty, R., Feldstein, M., and E., S., editors, *Handbook of Public Economics, vol. 5*, chapter 6, pages 329–390. Elsevier B.V.

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimal Income Taxation. *The Review of Economic Studies*, 38(2):175–208.

Piketty, T. and Saez, E. (2013). A Theory of Optimal Inheritance Taxation. *Econometrica*, 81(5):1851–1886.

Piketty, T. and Zucman, G. (2014). Capital is Back: Wealth-Income Rations in Rich Countries 1700–2010. *The Quarterly Journal of Economics*, 129(3, August 2014):1255–1310.

Riihelä, M., Sullström, R., and Tuomala, M. (2007). Varallisuuserot kasvussa. In Taimio, H., editor, *Talouskasvun hedelmät – Kuka sai ja kuka jäi ilman?* Työväen sivistysliitto, Helsinki.

Tenhunen, S. and Tuomala, M. (2010). On optimal lifetime redistribution policy. *Journal of Public Economic Theory*, 12(1):171–198.

4 TOP INCOMES AND INCOME DYNAMICS FROM A GENDER PERSPECTIVE: EVIDENCE FROM FINLAND 1995-2012

Terhi Ravaska, *Labour Institute for Economic Research, Helsinki, Finland*

Abstract

In this essay[1] *I study Finnish top incomes from a gender perspective using the Finnish register-based panel data over the period of 1995-2012. I find that that the under-representation of women at the top has been quite persistent in the overall top but the proportion of women in the top 1% has increased over 18 years. Women's wage share at the top has increased while the self-employment income has decreased. The top income females more often have an entrepreneurial background and are more often sharing a household with a high-income spouse.*

The gender-specific income distributions show that female incomes are less dispersed. In this study I also test whether top incomes can be assumed to be Pareto distributed. While the joint and men's top income distributions can be approximated with Pareto distribution throughout the observation period, the Pareto assumption gets more support for women after the year 2000. The female top income receivers have caught up with top earning men over time but I also show that females are more likely to move downwards from the top than men.

[1] I thank Ilpo Suoniemi and Matti Tuomala for the valuable comments as well as the seminar participants in STN-WIP 2017 seminar in Helsinki, EQINEC 2017 in New York, IIPF 2017 in Tokyo and Winter School in Inequality and Social Theory 2018 in Albi di Canazai. I also thank Joni Laukkarinen for the proofreading. Financial support from Palkansaajasäätiö is greatly acknowledged.

JEL: D31, J16, D63, D30
Keywords: income distribution, gender inequality, top incomes, income mobility

4.1 Introduction

The gender wage gap is a widely researched topic. There are much less studies on gender differences in the total incomes. Since the work by Piketty (2003) and Piketty and Saez (2003), inequality research has paid a lot of attention to the top incomes but little is known about the top income distribution from a gender perspective. While top incomes complement the traditional inequality analysis which focuses on the middle of the distribution and poverty, the gender perspective in the top incomes complements the analysis of overall gender inequality. This paper aims to bring these two perspectives together with empirical evidence from Finland. Finland performs relatively well in gender-equality (OECD, 2017) so this paper provides insight from a new perspective that can be useful for countries where similar steps towards gender-equality have not been taken.

Women at the top of the earnings distribution have been studied in many papers and from many perspectives (Albrecht et al., 2003; Guvenen et al., 2014; Bertrand et al., 2010; Fortin et al., 2017)[2] and this literature has summarized that the wage gap increases and women's presence decreases in the upper tail of the earnings distribution. A large part of the overall wage gap is also explained by the missing women at the top (Fortin et al., 2017). However, women and top incomes have been analysed only in few papers. Studies by Atkinson et al. (2018) and Roine et al. (2017)[3] show that women are under-represented at the top of the income distribution. The share of women decreases steadily further up in the distribution all over the world.

There are many explanations why the top of the income distribution has

[2]There is at least one paper also studying women in the wealth distribution by Edlund and Kopczuk (2009). This paper shows that big part of the wealth held by women is inherited in the US.

[3]Atkinson et al. (2018) study 8 countries which are Denmark, Norway, Spain, United Kingdom, Australia, Canada, Italy and New Zealand while Roine et al. (2017) provide evidence from Sweden.

so few women. For example, we know the gender wage gap is largely explained by the fact that on average women work in industries with lower wages and work fewer hours than men. Partly these observations can be attributed to chosen education paths. Lower earnings also lead to less saving opportunities for women and thus bigger gender based differences in capital income. However, if women's education and fields of work explain much of the gender gap, we should see more women at the right end of the income distribution over time as women have become more educated and shifted to work in the traditionally more male-occupied fields. Otherwise the persistent under-representation at the top of the income distribution may be due to a glass-ceiling effect[4].

In many international comparisons, Finland and the other Nordic countries, outperform in the women participation in the labour markets, education and politics and these countries are widely recognised as the most advanced countries in terms of gender equality at work (OECD, 2017). In Finland, women's share of the labour force has been over 45 % for three decades, the employment rates by gender are almost the same (in 2016 67,6 % compared to men's 69,8%[5]) and the education level is higher among women (Pietiläinen, 2013)[6]. How has this increase of educated women in the labour markets transformed the income distribution and its gender composition? The evidence from Sweden suggests that women have improved their representation at the top of the income distribution but still have more transitory incomes and more capital income than men (Roine et al., 2017). This study looks into the Finnish data to see whether Finland follows similar trends.

In this paper, I will study the Finnish top income distribution closely from a gender perspective. Firstly, I will explore the representation of women in the top income distribution, and study the income composition and background differences between men and women. I will show that there are clear gender differences between the top income receivers. At the very top, women tend to have larger capital income share than men and the share of wage income

[4]The term glass ceiling is used to define an unseen barrier that keeps women out of the top regardless of their qualifications.

[5]Statistics Finland Official Statistics on Employment.

[6]Women have completed more university degrees than men since 1985 (Pietiläinen, 2013, p. 18) and the gap has widened over time.

has only increased after the financial crisis. Compared to the men's income composition at the very top, the female share of capital income indicates that becoming rich by working is less common among women. This is verified by the fact that upper management positions are more common among male top income receivers while a large part of the women in top incomes have entrepreneurial backgrounds especially in the late 90s and early 00s.

Secondly, I will study the shape of the top income distribution by gender and show that the common distributional assumption of Pareto Type I does not hold. From the gender perspective we do not have any strong reason to assume that the male and female top incomes can be characterized with the same Pareto model. Even for the joint distribution the assumptions of Pareto distribution have rarely been questioned. The recent contribution by Jenkins (2017) shows that Pareto Type II model is more appropriate at least for the heavy tail in British income distribution. Instead of assuming that top incomes for both genders are Pareto distributed, I also tests this assumption for the Finnish top incomes. This can be seen as a contribution in its own right. In this part I will also compute the top income shares from the gender-specific distributions.

Thirdly, I will answer the question of how income mobility and income dynamics differ between genders. Annual cross-section measures do not give the full picture on income inequality. The income mobility between years contribute to the lifetime income differentials. For this reason I also extends the analysis by taking into account income mobility. From the top of the income distribution, individual can move only to the lower income groups and so the income mobility is measured as the persistence to stay in the top group over different periods. The question is, does this persistence differ between genders?

The analysis is based on the Finnish population's register data for the years 1995-2012. In the data the panel attrition occurs only due to death or emigration. Therefore long time periods of an individual's life can be observed. The panel structure of the data is used by extending the analysis of annual incomes to include average income for longer periods. The data is without top-coding so it is particularly well tailored to studying top incomes. There is a rich set of background variables included in the data from several official registers. The tax register data enables the decomposition of the sources of income. The tax unit in Finland is individual, however, the data includes a household identifier so family characters and spouse income can be used when studying the background of top income receivers. The main contribution of the study is to analyse top incomes from a gender perspective in detail with very extensive micro data.

The paper is structured in the following way: Section 4.2 introduces the data and income definitions. In section 4.3 I show time-series evidence on the overall top incomes and focus on the share of women in different top groups and how the incomes are composed. The subsection 4.3.3 discusses the background of the individuals in the top groups. Section 4.4 fits Pareto model separately for genders at the top of the income distribution. Section 4.5 presents the results with respect to income dynamics and gender. Section 4.6 concludes.

4.2 Data and income concepts

The data used in this study comes from the Statistics Finland's collection of administrative data for income distribution statistics. The dataset is constructed by taking a 10 percent representative sample of the Finnish population (approximately 500 000 individuals) and follows the individuals over time between years 1995-2012. Individuals exit the data if they emigrate or die. The individuals without address or who are institutionalized in any of the observation years are not included in the data.

The data includes a rich set of variables. The underlying register data originates from Population Register Center, Tax Administration, The Social Insurance Institution of Finland, National Institute for Health and Welfare, Finnish Centre for Pensions, the Register of Completed Education and Degrees and Financial Supervisory Authority. The data includes among other variables wage income, self-employment income[7], capital incomes (dividends and other capital income) and realized capital gains. The data also includes those with zero

[7]Self-employment income here refers to entrepreneurial income from agriculture, forestry or copyrights, and entrepreneurial income from business activity where the ownership is active on contrary to the passive owning of business which is taxed under capital income taxation rules.

incomes. The inflation is taken into account by deflating all income components to 2008 prices with Finnish consumer price index. The data also have a rich set of background variables, such as completed education, type of work, industry, day of birth and death, age in the end of the year, as well as information on the household type. There is no top-coding in the data.

In this study I concentrate on the adult population and thus exclude individuals below the age of 20. The tax unit in Finland is individual. There are some minor exemptions and subsidies that are family- or spouse based. An example of such an item is capital loss credit. The individuals with negative gross incomes are removed from the data (0.01% of the observations). The main analysis is complemented by also using three-year average incomes to reduce the effect of temporary income shocks. Here I have removed those observations that have gaps in the previous 3-year-periods (around 1,7 % of the observations over the years). I have winsorized the outlier observations, i.e individuals with very high incomes representing top 0.01% or higher, from the data. This is done due to the privacy restrictions set by the data provider but also to reduce the potential problems in interpretation of the results[8].

As the data is based on administrative records, it is more reliable than survey estimates of the top incomes. However, registers do not include all income sources. Such missing incomes are almost all interest income (which are taxed at source), some inter-household income transfers (for example child support is missing until 2010) and imputed rent for home-ownership. Also noticeable is that the income concept was updated in 2010 by including more accurate forestry income and child support income. For this reason the time-series before and after are not completely comparable. However, the effect is small. For example, the average equivalent income for the top decile was 0,5 % smaller than with the earlier income concept in 2013 (Statistics Finland).

The main income concept used is individual gross income excluding realized capital gains. The individual gross income is factor income with income transfers such as pensions or sickness benefits. The composition of these income items is shown in the appendix. The Finnish micro data on incomes is rich enough to build the series for both including and excluding realized capi-

[8]Winsorizing ensures that none of the data points is based on less than 30 individual observations.

tal gains. However, it is not clear if the realized capital gains are a good proxy for the accrued capital gains as tax changes affect the timing of selling assets as demonstrated in Burkhauser et al. (2015) and Armour et al. (2013). Also it has been noticed that top income shares are biased downward if accrued business income is not included (Alstadsæter et al., 2016).

Finland has a dual income tax system, where income from wealth (e.g. dividends, property rents and capital gains) are under capital income taxation and labour earnings or self-employment incomes are taxed under a progressive schedule. The capital income tax rate was flat until 2011 and since then there have been two tax brackets. The major tax reform in Finland, introducing the dual tax system occurred in the year 1993 so two years before the observation period starts in this study. There are smaller tax reforms possibly affecting the top incomes during the observation period and these are listed in the appendix.

Table 4.A1 in the appendix shows descriptives statistics separately for each year and gender. From these summary statistics we can conclude a few points about the changes in the overall income distribution over the years. The gross income distribution has become more dispersed for both genders but more so for men. Also the average gross income has increased for women during the observation period but for men the financial crisis lowered the average income. Also over time the mean absolute income gap between men and women has increased. The average size of capital income and capital gains have increased for women since the 90s.

4.3 Top incomes and women between 1995-2012

This section starts with the overall review of the top income shares in Finland over time. This step is taken in order to place the gender-specific analysis into context. After this the female representation and income composition at the wider top are analysed. In the third subsection simple probability models are run in order to determine what kind backgrounds women and men at the top of the income distribution have.

4.3.1 Trends in overall top incomes

Trends in the top incomes over the world, including Finland, is summarized in Atkinson et al. (2011). In international comparison Finnish top income shares are low but differ from continental Europe by showing a clearer upward since the mid 90s. The Finnish top incomes have been studied by Jäntti et al. (2010) in more detail until the year 2004. The stark finding is that the top incomes increased rapidly in the late 90s. Here I extend this analysis to the year 2012 by calculating the income shares based on the individual income and conclude that the top income shares are still at similar levels as in the beginning of the 2000s.

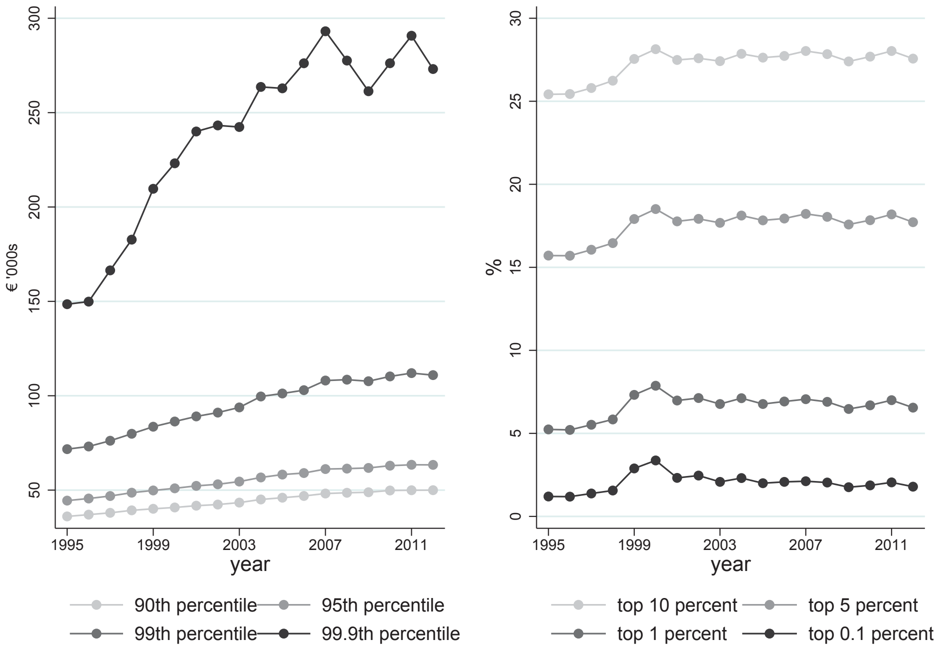
The median individual income without realized capital gains was approximately 17,600 € in 1995. 18 years later it had increased 35 percent to 23,700 €. The median income in the top 10% group grew around 43 percent from the 44,000 € in 1995. In the top 1% and top 0.1% the income growth was 62 and 85 percent in the same period, respectively. The faster income growth in the top incomes has widened the income distribution.

Figure 4.1 shows the lower thresholds and income shares for selected top income groups based on the individual income[9]. The top 10% income threshold was approximately 36,000 € and the threshold 18 years later approximately 50,000 €. To be included in the top 1% one needed to have over 71,800 € of income in the beginning of the observation period compared to 111,000 € in 2012. This translates to 54 % higher income requirement in 2012 in order to be in the top 1%. In the very top, above 99.9 percentile, the income requirement grew even more, nearly 85 percent. The income shares for these groups increased rapidly in the late 90s but in the 00s there has been little changes. The top 1% received 5 percent of the total income in 1995 and approximately 7 % in the 2012.

About 75 percent of the income of the top 10% is wage income (figure 4.2). Over the years the share of capital income has increased from 3 % to 8 % but

[9] Inequality is preferably studied with equivalised household income but since I am studying top incomes by gender, the individual income is used. While the trend qualitatively is the same whether the top income shares are measured by individual income or equivalised household income, the levels differ. Hence note that the figures here are not comparable to Jäntti et al. (2010).

Figure 4.1 The annual income thresholds for incomes above different income percentiles (left panel) and income shares (right panel).

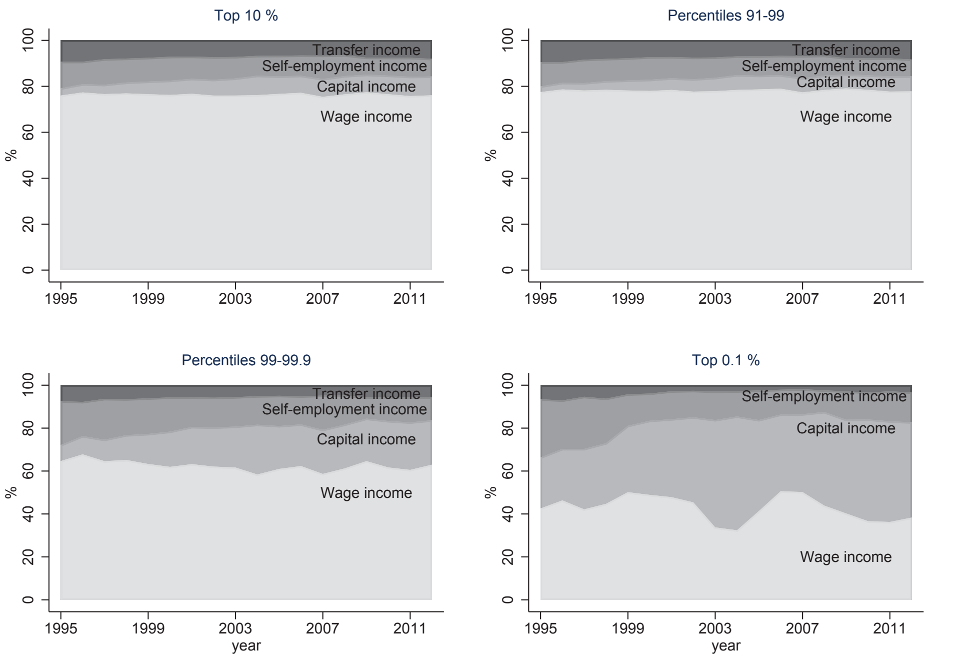


Notes: Income measure is gross individual income excluding realized capital gains. Incomes deflated to 2008 euros.

most visible is the important role of capital in the top 1% or higher groups. In these groups it is also clear that the role of capital became more important in the end of the 90s and early 00s. At the very top, the self-employment income was replaced by the capital income. An explanation for this is the income shifting caused by the tax reform in 1993 which created incentives for entrepreneurs to report their income as capital income rather than self-employment income (Selin and Pirttilä, 2011).

The years when the share of capital income grew corresponded with the years when the top income shares grew the fastest. On average 38 % of total income is capital income in the top 0.1% and during the years of high stock market returns capital share is as high as half of the income (excluding realized capital gains).

Figure 4.2 The annual income composition for selected top income groups

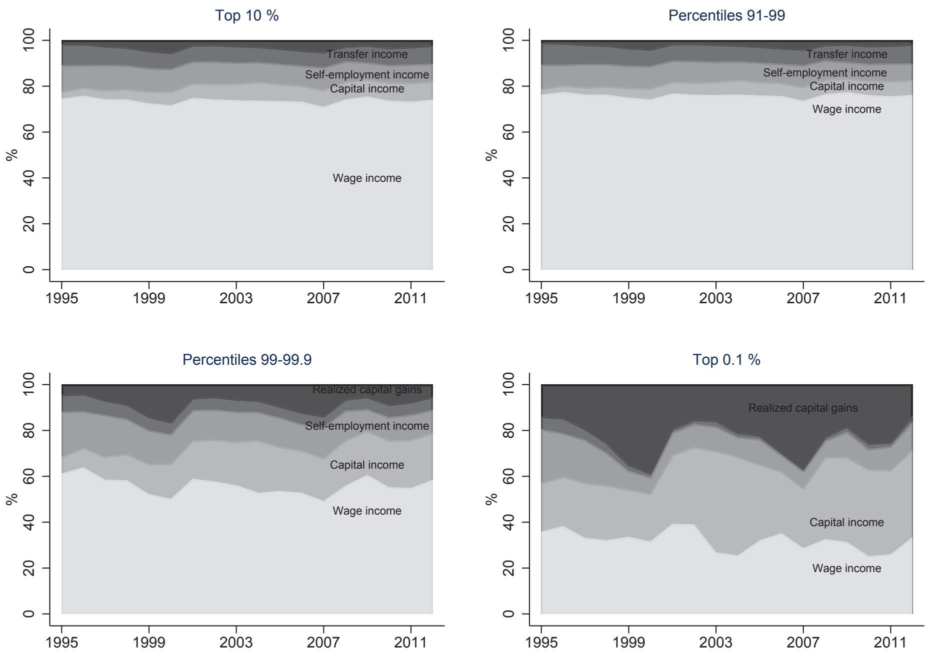


Notes: Income measure is gross individual income excluding realized capital gains.

Just as the annual income from the capital is concentrated to the top, also the realized capital gains are targeted to the top. Over the years, 11 to 14 percent of all realized capital gains are received by the richest 1%. Figure 4.3 presents the income composition with realized capital gains. In the groups below 99th percentile, the capital gains have relatively little influence on total income during the observation period as income is composed mainly by the wage income. However, the pattern is totally different in the high income groups where beside annual income stemming from wealth the active selling of assets is important. In the top 0.1% these realized capital gains are on average 22 percent of total income and during the stock market booms the capital gains share is as high as 38 %. In the rest of the top 1% the share of realized capital gains varies from 4 to 16 percent.

The observation period includes periods of strong economic growth but

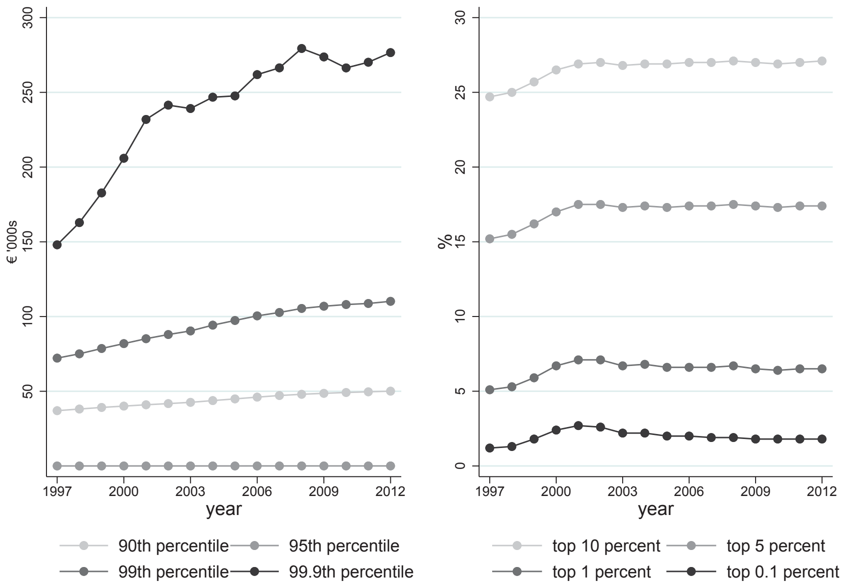
Figure 4.3 The annual income composition for selected top groups.



Notes: Income measure is gross individual income including realized capital gains.

also deep economic crisis and subsequent recession. To reduce the annual volatility in income, the figure 4.4 shows the evolution in top income thresholds and income shares using individuals' average income from the previous three years. This reduces the height of the spikes in the data but the overall trend remains: in the end of the 90s the incomes of the top increased rapidly whether measured as income shares or looking at the thresholds for getting in to the top group. After this period there has been very little movement in either direction. The income shares are 2-3 percentage point higher than in the beginning of the observation period.

Figure 4.4 The income thresholds for 3-year average incomes above different income percentiles (left panel) and income shares (right panel).



Notes: Income measure is average gross individual income during the previous 3-year excluding realized capital gains. Incomes deflated to 2008 euros.

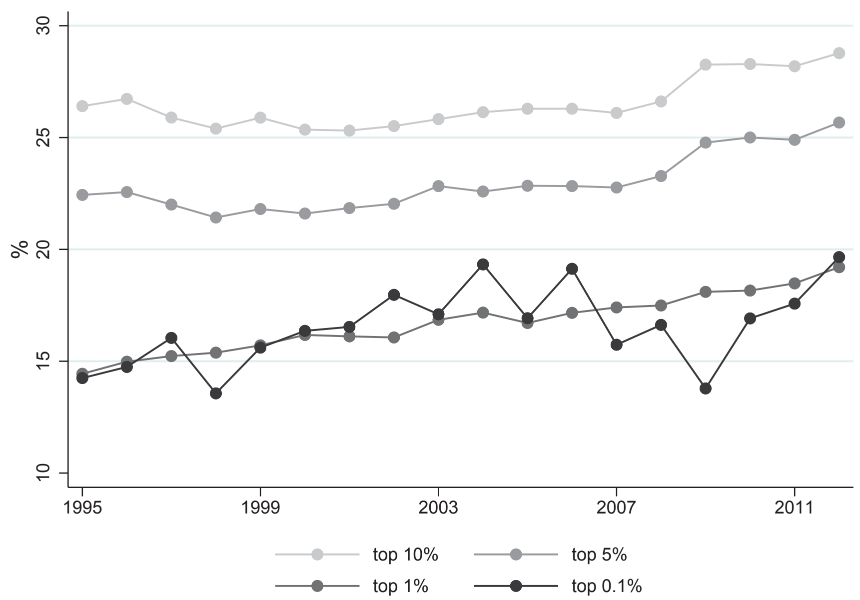
4.3.2 Share of women at the top and their income composition

This section will focus on the share of women in top incomes. During the observation period the trends in the labour market between genders have been similar so the observed differences between men and women are not stemming from the increasing attachment of women to labour markets. The rapid growth in the female participation in the labour markets already happened in the 70s and 80s.

Figure 4.5 shows the share of women in different groups based on total income excluding the realized capital gains[10] (figure 4.A1 is appendix shows

[10]Figure 4.A2 in appendix shows the share with the income including the realized capital gains. The realized capital gains do not affect the top 5% or 10% groups. However, women share increase in the higher groups during a stock markets peaks. For example, before the IT bubble, in year 2000 the top 0.1% had 22,5 % of women. This is natural as women in these higher groups tend to have more capital income.

Figure 4.5 Share of women in different income groups, years 1995-2012



Notes: Income distribution based on gross individual income excluding realized capital gains.

the share with 3 year average income). In the broader top 10% and top 5% groups the share was decreasing slightly during the rapid economic growth in the late 90s. Overall the growth in the share of women has been quite flat but there is a clear jump, approximately 2 percentage points, in the aftermath of the financial crisis. The share of women in the top 10% group was almost 29% at the end of 2012 which was an increase of 4 percentage points from the period's lowest value. In the top 5% one in four have been women since 2009.

Looking at the top 1%, the share of women has increased steadily throughout the period from less than 15 percent to almost 20 percent. Together with the fact that the share of women in the overall top 10% was stagnant much of the period, this means that the women within the top 10% have become richer. In fact, during the observation period, the mean incomes in the top 1% percent grew 64,1 % for women and 74,7 % for men and in top 10% 50,2 and 62,0 respectively (table 4.1). Looking more closely at the yearly figures, we can also notice that even though there are fewer women at the very top, these women have incomes that compare to men's. In the top 1% women have higher mean incomes for half of the observation years.

Table 4.1 Mean and median incomes (excluding realized capital gains) in top 10% and top 1% groups

year	Women				Men			
	median top 10%	mean top 10%	median top 1%	mean top 1%	median top 10%	mean top 10%	median top 1%	mean top 1%
1995	42 865.42	48 486.34	86 925.66	113 331.02	45 202.79	53 226.71	88 814.97	106 080.42
1996	43 860.08	49 501.97	89 211.32	112 535.49	46 344.94	54 521.82	90 716.10	108 185.36
1997	45 093.70	51 760.26	91 706.51	126 434.39	47 601.29	56 640.98	95 206.24	116 999.53
1998	46 743.75	53 847.42	97 389.35	130 917.24	49 420.13	59 489.32	101 165.81	128 891.79
1999	47 741.80	56 484.97	104 137.18	154 619.58	50 676.37	65 643.20	107 009.79	170 666.13
2000	48 835.91	58 924.01	107 041.02	169 315.67	51 822.78	68 418.53	110 748.82	187 668.23
2001	501 66.36	60 055.17	112 733.27	167 769.02	53 079.75	67 198.46	115 454.17	165 570.32
2002	51 048.28	61 665.55	113 692.44	177 764.67	53 928.85	68 786.91	117 357.97	172 137.60
2003	52 605.38	64 043.94	115 822.79	186 088.53	55 274.43	69 055.52	121 039.45	163 427.13
2004	54 415.07	67 231.12	127 687.14	203 540.26	57 640.75	72 933.66	128 523.52	178 333.82
2005	55 892.89	67 323.49	126 076.20	187 636.40	59 238.39	73 795.46	131 260.77	174 347.21
2006	56 723.49	68 180.99	128 271.76	183 619.09	60 030.33	75 717.40	135 291.06	183 987.29
2007	58 608.85	70 179.33	135 311.30	183 349.82	62 174.17	78 906.68	142 776.16	195 163.74
2008	59 091.62	70 551.68	135 793.30	182 043.61	62 387.21	78 919.01	141 582.80	191 658.16
2009	59 312.60	69 569.15	130 520.15	168 291.67	62 874.39	78 274.82	139 372.16	181 421.90
2010	60 713.94	72 151.46	136 847.48	187 694.69	63 990.84	80 148.82	143 076.29	188 163.93
2011	61 186.55	73 423.96	138 539.73	196 610.29	64 458.88	81 571.80	147 903.87	198 195.63
2012	61 267.30	72 841.05	137 933.30	186 011.96	64 380.55	80 052.94	143 935.06	185 279.08
growth % 1995-2012	42.93	50.23	58.68	64.13	42.43	50.40	62.06	74.66

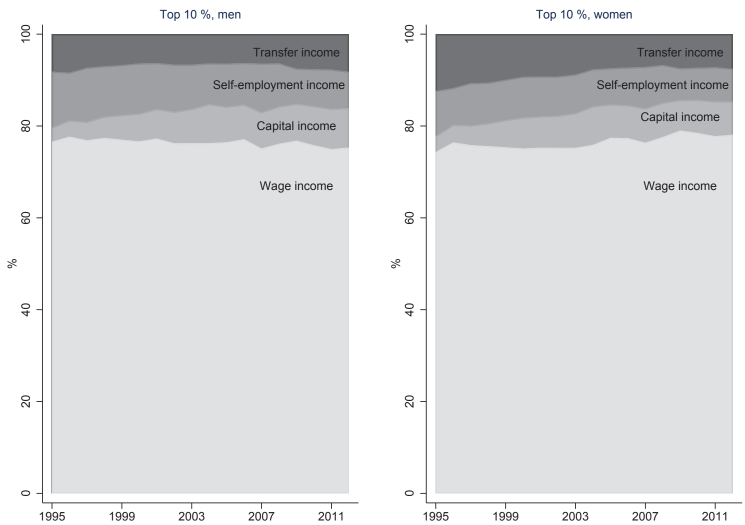
Note: Incomes deflated to 2008 euros.

The figures 4.6, 4.7 and 4.8 show the income composition in different top income groups for men and women. While it is remarkable that wage share is similar in the top 10% between men and women (around 75%), the higher income group show clear differences, where women have less wage income but more capital or self-employment income. In the top 10% group, women's share of transfer income has decreased over the years. The transfer income in these income groups mainly consists of pension income and secondly disability benefits for men and family and survivor's benefits for women. The reduction in the share of survivor benefits explain much of the decrease in the overall transfer income.

The lower share of wage income in the higher income group could indicate that becoming very rich by working for an outside firm is less common among women. The self-employment income in the top 1% is more important for the women. The female entrepreneurship indeed is more common in the top groups, however the share of female entrepreneurs among top 1% has decreased while at the same time the wage share has increased towards the end of the observation period. In 1995 within the top 1% females, approximately one in four were self-employed while among men 14 percent were self-employed. The male entrepreneurship at the top has increased around 4 percentage point while within women the share is almost the same. Despite these trends the self-employment income has decreased in total. Partly this could stem from the income shifting that the 1993 tax reform induced (Selin and Pirttilä, 2011).

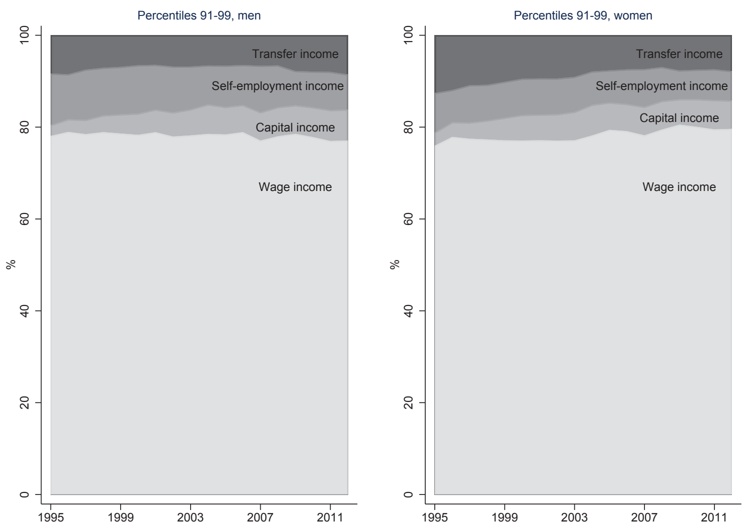
The most common socioeonomic status for men at the top was senior official and upper management. However, the females have taken over more upper managers positions in the 2000s. At the end of the period there was almost the same share (around 30%) of upper managers among women and men top income receivers. This is also supported by the previous observation that the wage income share has increased for women.

Figure 4.6 Income composition in top 10%



Notes: Income measure is gross individual income excluding realized capital gains.

Figure 4.7 Income composition in 91-99 percentiles



Notes: Income measure is gross individual income excluding realized capital gains.

Figure 4.8 Income composition in 99-99.9th percentiles

Notes: Income measure is gross individual income excluding realized capital gains.

Figure 4.9 presents the top 0.1%. There is a dramatic change in the income composition and a clear gender difference compared to the rest of the top 1%. The wage share for men range between 38 and 55 percent while women's wage share is on average 22 % over the years observed. The share of self-employment income is approximately 10-15 percent higher among women, and the same is true for the capital income. Unfortunately the data does not include inheritance information, so I cannot determine the role of bequests for the capital share. In conclusion, it seems that the men in the overall top seem to earn from working for somebody else, while women at the top either earn by owning a company or get high returns from owning assets. The higher share of capital also translate as a higher representation of women at the very top of the income distribution including realized capital gains (figure 4.A2 in the appendix).

Figure 4.9 Income composition in top 0.1%

Notes: Income measure is gross individual income excluding realized capital gains.

4.3.3 Who becomes a top income receiver?

In this section I study the background characteristics of the top income receivers. I estimate a logistic regression where the probability of being a top income receiver is regressed separately for women and men on their own and on their spouse characteristics. The regression equation is written as:

$$D_i = \alpha + \sum \beta_o X_i^{own} + \sum \beta_s X_i^{spouse} + \theta_s Y_i^d + \epsilon_i, \tag{4.1}$$

where D_i is a dummy for being in the top income receiver group (top 10% group in joint distribution)[11] and X_i contains a vector with the background variables for the individual herself and for the spouse if a spouse exists. I am especially interested in which fields of work and education levels are associated with a top income position and if there is an association between the spouse's income decile and women's probability to be in the top group. The background characteristics are education, field of work, number of children, marital status, mother tongue and region of residence. I also include spouse income decile Y_i^d in one of the regression specifications. The dependent variable is equal to 1 if the individual belongs to the top group in any year, and zero otherwise and the parameters to be estimated are α, β_o, β_s and θ_s. The industry and region variables are classified according to the TOL2008[12] and education level is categorized in 6 categories. The regressions control also for year fixed effects, age and age square effects. The mother tongue variable is recoded as Finnish speaking, Swedish speaking and other languages.

Table 4.2 shows the main estimated marginal effects from each regression specification. The full list of variables is shown in appendix 4.6. The first and second column show the association between different characteristics and being in the top 10% for single-adult households separately for men and women[13]. From the fields of work, finance sector is most strongly associated with being in the top 10%. For women the legal sector also increases the probability to be in the top 10%. If an individual has studied in the STEM field[14],

[11] I have also estimated the model for top 1% but as this smaller group is more prone to small sample bias in this logistic regression, I report here only the results for top 10%.

[12] http://www.stat.fi/meta/luokitukset/toimiala/001-2008/index_en.html

[13] The pooled sample estimates available upon request from the author.

[14] STEM=Science, technology, engineering and mathematics.

this is positively associated with high incomes but the marginal effect is small. Education level consistently increases the probability to be in the top. For the singles subsample the female entrepreneurship is not a strong correlate with being in the top 10% and for men it is even negative. However, entrepreneurship is associated positively and more strongly among the women who have a spouse (columns 3 and 5).

Columns 3 to 6 present the estimated marginal effects for the subsample of cohabiting individuals. For person's own characteristics the coefficients are for the most part in line with the singles subsample. The last rows of the table also show the marginal effects for the spouse characteristics. The last rows show the estimates for association between being a top income receiver and the spouse's income decile. For women these coefficients are also reproduced conditional on woman's own education level in figure 4.10. We see that there is a strong positive association between having a spouse in the highest decile and being in the top 10%. For women, this increases the probability by 3.9 percentage points and for men 5.1 percentage points (compared to the baseline where the spouse is in lowest income decile).

The potential explanation for the strong positive correlation between a spouse's income decile and the probability to be in the top are assortative mating (meaning the positive relationship between the couple's income ranks or education already before forming a joint household) or income shifting between spouses (especially through firm and asset ownership). The income shifting from husband to wife gets less support from the estimated marginal effects as the association decreases and gets close to zero if spouse is in the top 1%. With income shifting between spouses, the expected association should be increasing in spouse's income. There is also an asymmetry in the marginal effects. If the woman is in the top 1%, the probability for the man to be in top 10% increases by 5.6 percentage points while women's probability increases by 0.5 percentage points.

In the current study the assortative mating hypothesis cannot be adequately tested because the data does not show the spouse's characteristics before the joint household is formed. However, using the years before couple gets married, there is some support for assortative mating. 34 % of individuals who are in the top decile (measured with previous 3-year average income) of the

Table 4.2 Marginal effects from logistic regression, dependent variable being in the top 10%

VARIABLES	Women (single)	Men (single)	Women (cohabiting)	Men (cohabiting)	Women (cohabiting)	Men (cohabiting)
Working as professional in						
finance	0.0533***	0.0656***	0.0599***	0.122***	0.0629***	0.127***
	(0.00457)	(0.00814)	(0.00338)	(0.00793)	(0.00385)	(0.00827)
legal services	0.0187***	-0.00390	0.0225***	-0.0106***	0.0219***	-0.0100***
	(0.00274)	(0.00254)	(0.00217)	(0.00265)	(0.00243)	(0.00290)
health services	0.0166***	0.0189***	0.0122***	0.0368***	0.0146***	0.0391***
	(0.00247)	(0.00588)	(0.00182)	(0.00591)	(0.00213)	(0.00627)
Education level						
secondary level	0.0123***	0.0135***	0.00989***	0.0439***	0.00607***	0.0345***
	(0.00180)	(0.00297)	(0.00140)	(0.00317)	(0.00168)	(0.00368)
lowest level tertiary	0.0248***	0.0364***	0.0266***	0.0731***	0.0243***	0.0634***
	(0.00155)	(0.00254)	(0.00123)	(0.00256)	(0.00145)	(0.00287)
lower-degree level tertiary	0.0400***	0.0680***	0.0498***	0.140***	0.0466***	0.126***
	(0.00225)	(0.00341)	(0.00187)	(0.00333)	(0.00205)	(0.00358)
higher-degree level tertiary	0.115***	0.122***	0.141***	0.250***	0.137***	0.227***
	(0.00343)	(0.00454)	(0.00269)	(0.00430)	(0.00288)	(0.00461)
doctorate or equivalent	0.205***	0.192***	0.255***	0.386***	0.251***	0.359***
	(0.0122)	(0.0132)	(0.00858)	(0.00986)	(0.00866)	(0.0101)
education in STEM	0.0116***	0.0116***	0.0223***	0.0211***	0.0268***	0.0268***
	(0.00298)	(0.00299)	(0.00245)	(0.00316)	(0.00281)	(0.00341)
Occupation						
self-employed	0.00206	-0.0201**	0.0382***	0.0130**	0.0405***	0.0294***
	(0.0159)	(0.00854)	(0.00625)	(0.00616)	(0.00698)	(0.00715)
upper management	0.187***	0.270***	0.165***	0.384***	0.172***	0.405***
	(0.0176)	(0.0115)	(0.00736)	(0.00706)	(0.00803)	(0.00799)
senior employees in R&D	0.0119	0.0379***	0.0142**	0.119***	0.0289***	0.152***
	(0.0160)	(0.00931)	(0.00617)	(0.00669)	(0.00696)	(0.00767)
senior employees in education	-0.0267*	-0.0618***	-0.0239***	-0.0673***	-0.00872	-0.0364***
	(0.0154)	(0.00917)	(0.00565)	(0.00673)	(0.00646)	(0.00777)
Other senior employees	0.0448***	0.0579***	0.0537***	0.172***	0.0696***	0.199***
	(0.0158)	(0.00997)	(0.00611)	(0.00741)	(0.00690)	(0.00831)
supervisors	-0.0167	-0.00170	-0.00578	0.0217***	0.0114	0.0503***
	(0.0160)	(0.00887)	(0.00613)	(0.00632)	(0.00703)	(0.00740)
clerical workers, independent	-0.0647***	-0.0530***	-0.0503***	0.0252***	-0.0345***	0.0526***
	(0.0152)	(0.00851)	(0.00549)	(0.00649)	(0.00631)	(0.00751)
clerical workers, routine	-0.0917***	-0.112***	-0.0628***	-0.0794***	-0.0503***	-0.0528***
	(0.0155)	(0.0103)	(0.00591)	(0.0111)	(0.00683)	(0.0122)
lower-level admin. & clerical occ.	-0.0930***	-0.0797***	-0.0792***	-0.0930***	-0.0682***	-0.0672***
	(0.0152)	(0.00833)	(0.00542)	(0.00612)	(0.00623)	(0.00716)
workers in agriculture	-0.101***	-0.150***	-0.0914***	-0.165***	-0.0801***	-0.132***
	(0.0171)	(0.00878)	(0.00614)	(0.00992)	(0.00761)	(0.0114)
manufacturing workers	-0.0647***	-0.0572***	-0.0562***	-0.0559***	-0.0375***	-0.0245***
	(0.0165)	(0.00808)	(0.00631)	(0.00597)	(0.00758)	(0.00714)
other production workers	-0.101***	-0.121***	-0.0851***	-0.146***	-0.0732***	-0.117***
	(0.0155)	(0.00828)	(0.00577)	(0.00649)	(0.00675)	(0.00770)
distribution and service workers	-0.107***	-0.132***	-0.0837***	-0.162***	-0.0698***	-0.134***
	(0.0152)	(0.00800)	(0.00549)	(0.00621)	(0.00637)	(0.00740)

Table 4.2: Marginal effects from logistic regression, dependent variable being in the top 10%, cont'd

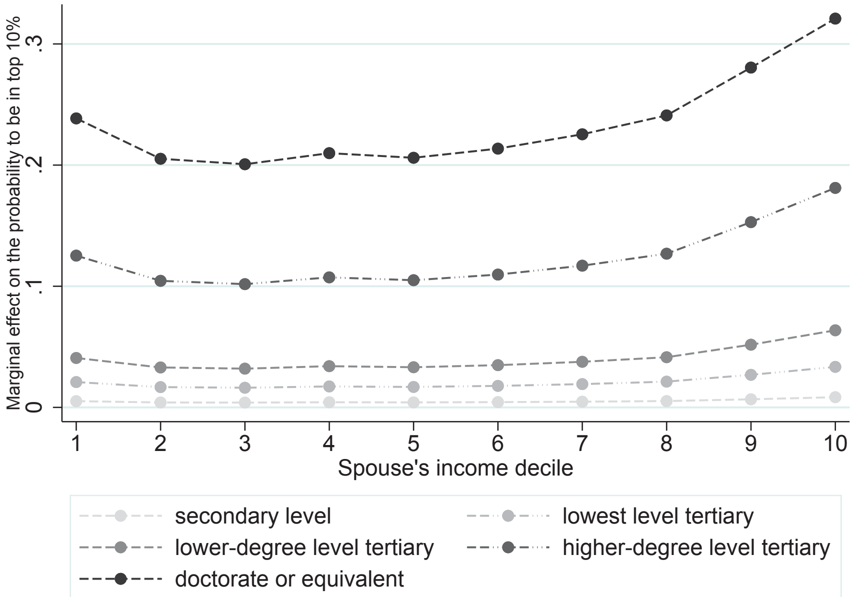
VARIABLES	Women (single)	Men (single)	Women (cohabiting)	Men (cohabiting)	Women (cohabiting)	Men (cohabiting)
spouse characteristics						
self-employed					0.00418	-0.0145*
					(0.00447)	(0.00762)
upper management					-0.0243***	-0.0380***
					(0.00443)	(0.00812)
senior employees in R& D					-0.0230***	-0.0431***
					(0.00441)	(0.00769)
senior employees in education					-0.0210***	-0.0480***
					(0.00467)	(0.00753)
other senior employees					-0.0196***	-0.0391***
					(0.00448)	(0.00753)
education in STEM					-0.00653***	-0.00297
					(0.00186)	(0.00507)
spouse income group						
2nd decile					-0.0133***	-0.0317***
					(0.00270)	(0.00326)
3rd decile					-0.0150***	-0.0348***
					(0.00269)	(0.00333)
4th decile					-0.0115***	-0.0334***
					(0.00268)	(0.00336)
5th decile					-0.0130***	-0.0307***
					(0.00268)	(0.00343)
6th decile					-0.0101***	-0.0247***
					(0.00266)	(0.00347)
7th decile					-0.00538**	-0.0201***
					(0.00263)	(0.00355)
8th decile					0.00103	-0.00445
					(0.00260)	(0.00369)
9th decile					0.0186***	0.0186***
					(0.00265)	(0.00394)
10th decile					0.0392***	0.0517***
					(0.00280)	(0.00453)
top 1%					0.00534**	0.0538***
					(0.00269)	(0.0108)
Observations	1,017,288	964,654	1,017,288	964,654	812,841	810,482
Sample mean prob. to be in top 10%	0.0414	0.1159	0.0760	0.2448	0.0860	0.2656

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note: All predictors at their mean values. Year fixed effects also included in the regression.

income distribution one year before marriage marry a person who is either in the 9th or 10th decile. This is especially strong for women: 65 percent of women in the top decile marry either from the same or 9th income decile (while the same is true for 25% percent for men).

Figure 4.10 Association between spouse's income decile and probability of being in the top 10% for women, conditional on own education.



Notes: The figure displays the marginal effect on the baseline probability to be in the top 10% conditional on spouse's income decile. The coefficients are based on the last two columns in the table 4.2.

4.4 Gender-specific income distributions

While section 4.3 explored women at the top of the joint income distribution, this section analyses the gender-specific income distributions. By analysing the gender specific income distribution we can explore whether the inequality process is similar between genders and do the gender specific distribution become similar over time. First I present the income shares calculated from the gender-specific distributions which tell about the income concentration within each gender. After this I explore the shape of the gender-specific top income distributions.

4.4.1 Top income shares 1995-2012

Table 4.3 shows the income shares from the gender specific income distribution in the top decile. Looking at the gender differences in the top decile, we notice that income shares in different years are 3.5 to 6 percentage points higher within men's distribution compared to women's. While the top decile collect on average 24.7 percent of total income in the female distribution over the observation years, the top in the male distribution collect 29.4 percent. That is, the top income distribution within men is more unequal.

Relative to the men's top income shares, the changes in the income concentration within the top of the women's distribution has been modest. The increased inequality at the end of the 90s also shows up in the gender specific distributions; there was a clear increase in the gender specific income shares in the year 1999 and after, whether measured in absolute or relative terms. However, the increase in inequality was much higher in the men's distribution. Income shares have been more volatile for men and have ranged between 26,5 % and 31 %. During the years of strong economic growth (and high stock market returns) the gap between women's and men's income shares in the top 10% got larger.

A closer look within the top income decile reveals that much of the increase or decrease in inequality between the years comes from the very rich gaining or losing more over the years. There is extremely small movement in the income share within the percentiles 91-99[15]. The movements in the income shares of the top 0.1% and percentiles 99-99.9 explain much of the overall developments within the top decile. For example between 1998 and 1999 the women's top 10% income share increased 1,21 percentage point while there was practically no change in the income share when the top 1% is excluded. The same is true in the men's distribution. The share of income for the top excluding the top 1% actually shrank even while there was almost a 3 percentage point increase in the income share for the top 10%.

Together with the observations on the overall top income shares presented in section 4.3.1 we can conclude that the growth in inequality especially during the late 90s was driven in large part by the very rich men gaining more

[15] Tables available upon request from the author.

from the economic growth. Beside growing inequality within the top 10%, at the end of the 90s the gender differences in the income shares also got larger. The economic growth periods have been more favourable for men. However, this is not an indication of a more favourable labour market and saving opportunities for men. One needs to bear in mind that the incomes within the top groups of the men's distribution are higher than in the women's income distribution and thus the economic growth and downturn periods affect men and women disproportionally. The next section explores the shapes of the gender-specific tail distributions more closely.

Table 4.3 Top income shares from gender-specific income distribution

	Women				Men			
year	top 10%	top 5%	top 1%	top 0.1%	top 10%	top 5%	top 1%	top 0.1%
1995	23,03 %	13,87 %	4,43 %	1,13 %	26,59 %	16,83 %	6,00 %	1,63 %
1996	23,07 %	13,82 %	4,31 %	1,00 %	26,63 %	16,88 %	6,09 %	1,74 %
1997	23,49 %	14,24 %	4,66 %	1,23 %	27,06 %	17,32 %	6,40 %	1,80 %
1998	23,87 %	14,60 %	4,92 %	1,34 %	27,97 %	18,27 %	7,32 %	2,39 %
1999	25,08 %	15,87 %	6,14 %	2,32 %	30,84 %	21,38 %	10,43 %	4,99 %
2000	25,72 %	16,51 %	6,67 %	2,65 %	31,81 %	22,40 %	11,40 %	5,80 %
2001	24,43 %	15,17 %	5,39 %	1,66 %	29,53 %	19,93 %	8,89 %	3,47 %
2002	24,46 %	15,26 %	5,49 %	1,75 %	29,45 %	19,85 %	8,78 %	3,44 %
2003	24,77 %	15,55 %	5,67 %	1,84 %	28,89 %	19,14 %	7,85 %	2,50 %
2004	25,12 %	15,88 %	5,98 %	2,06 %	29,86 %	20,12 %	8,70 %	3,10 %
2005	25,22 %	15,96 %	5,95 %	2,03 %	29,83 %	20,00 %	8,47 %	2,85 %
2006	25,58 %	16,30 %	6,28 %	2,27 %	30,87 %	21,19 %	9,72 %	3,89 %
2007	25,48 %	16,14 %	5,96 %	1,78 %	31,38 %	21,68 %	10,02 %	3,73 %
2008	25,16 %	15,85 %	5,79 %	1,83 %	29,94 %	20,18 %	8,67 %	3,06 %
2009	24,60 %	15,30 %	5,28 %	1,43 %	29,28 %	19,38 %	7,81 %	2,40 %
2010	25,24 %	15,91 %	5,81 %	1,82 %	30,07 %	20,19 %	8,48 %	2,84 %
2011	25,42 %	16,10 %	5,94 %	1,89 %	31,02 %	21,26 %	9,68 %	3,99 %
2012	24,88 %	15,53 %	5,40 %	1,43 %	29,14 %	19,21 %	7,54 %	2,19 %

Notes: income measure is individual gross income excluding realized capital gains.

4.4.2 Pareto model for gender-specific income distributions

In modelling top incomes, a typical practice is to assume a Paretian distribution. In a recent contribution by Atkinson et al. (2018), the Pareto Type I assumption was used in order to estimate Pareto α parameters separately for men and women. The differences in αs were defined as sort of a "glass ceiling" because with the Pareto curve estimation one can show how fast, compared to men, the women disappear from the top. However, even when a Pareto model might be a good approximation for the joint distribution, there might be differences when we are looking at the gender-specific top income distributions. Testing for the Pareto distribution hypothesis is often neglected even when the estimation of the parameters is meaningful only if the used data is drawn from the same distribution.

In the case of top incomes characterized with a Pareto model[16], the complementary CDF (survivor function) is in the form

$$S(y) = (\frac{y}{y_m})^{-\alpha}, \text{ when } y > y_m. \tag{4.2}$$

In the notation y denotes the income, $y_m > 0$ is the threshold where Pareto assumption is valid and the parameter α is the shape parameter which indicates how heavy the top tail in the distribution is. The smaller the α, the more heavier top tail the distribution has. The α needs to be greater than 1 in order to have a finite mean. In this case the mean is $\frac{\alpha y_m}{\alpha - 1}$. The α parameter is easily estimated if the lower threshold y_m is known with OLS or using maximum likelihood (ML) estimator.

Taking logarithm from each side of equation 4.2 results in
$log\ S(y) = \alpha\ log\ y - C$, where C is a constant. We notice that the relationship between the complementary CDF and income is linear in a log-log plot. The usual first test to see if data is Pareto distributed is to graph this relationship. Left panel of the figure 4.11 shows the log-lot plot and the linear fit for the top 5 % for selected years[17]. In the beginning of the observation period the fit is worse in the women's distribution (lower R-squared), but over time as

[16] Pareto model here refers to Pareto type I model.

[17] Other years available in the supplementary material at `http://www.labour.fi/?wpfb_dl=4649`.

women have catched up with men, the fit improves. Also the absolute slope parameter for women is higher indicating that the upper tail is less concentrated. However, the log-log plots and the linear fits are only necessary conditions and are not sufficient in their own (Clauset et al., 2009; Cirillo, 2013). This type of analysis should be complemented with other graphical tools and distributional tests.

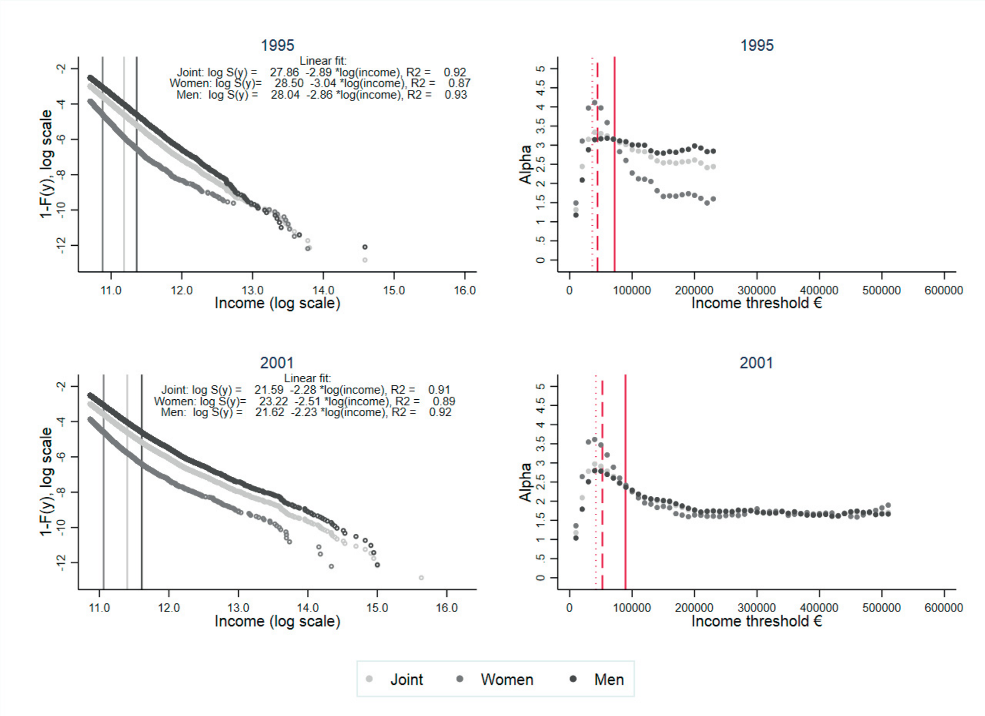
One property of Pareto distribution is that if the minimum threshold is correct, the estimated α parameter should be stable above the threshold[18]. The right panel of figure 4.11 plots the estimated α against the minimum thresholds[19]. The vertical lines give the threshold for top 10%, 5% and 1% in the joint distribution. The figure reveals that at the turn of the century and after, the Pareto tail is more prevalent. Due to the smaller number of women at the top the men's distribution follows the joint distribution. In previous applications the common practice has been to assume the Pareto tail to be a valid approximation for incomes above the 95th or 99th percentile, however figure 4.11 reveals that especially for the female distribution, the threshold is higher than the 99th percentile. Kolmogorov-Smirnov goodness-of-fit test statistics indicate that the estimated Pareto distributions fits the data best when the lower threshold is set approximately to 200 000 euros for years after 2000 and this corresponds to 99.5th percentiles and above. However, the number of observations are small at the very top and so this result needs to be replicated with the full population data in the future.

The estimation of alphas for years 1995-1999 break down in the upper tail of the women's distribution. This is because either there are too few observations or the tail indeed is not Pareto distributed for women. It is hard to determine in a similar way to Atkinson et al. (2018) whether the "glass ceiling" has got thinner over time in Finland. Comparing the most stable alpha estimates show that in some years the female alpha has been below men's while for other years the opposite is true. However, from the overall analysis we can conclude that women have caught up with men and in this sense the "glass ceiling" has got thinner.

[18]The ratio $\frac{y^*}{y_m}$ where the y^* is the average income above the threshold y_m is constant for all y_m in Paretian distribution. The ratio equals $\frac{\alpha}{1-\alpha}$ and thus the α is constant.

[19]α parameter is estimated by maximum likelihood (Stata package paretofit). The method of maximum likelihood gives unbiased parameter estimates in the limit of large sample size.

Figure 4.11 Fitting a Pareto distribution to the top of the income distribution: log-log plot (left panel) and estimated alphas (right panel)

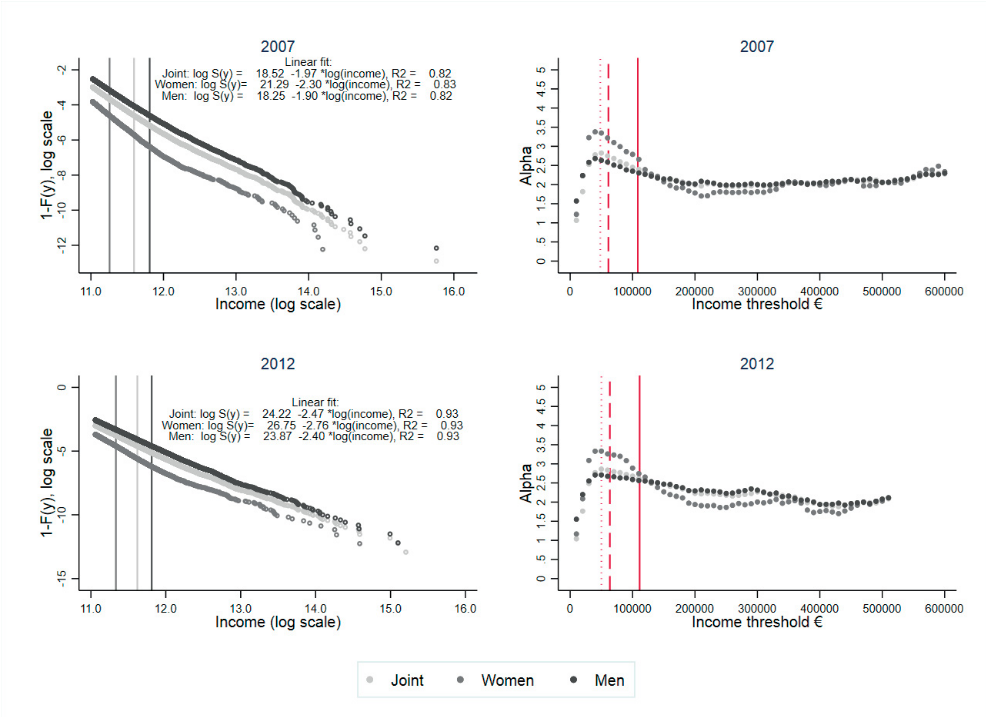


4.5 Income dynamics at the top by gender

There is relatively little research about the top income mobility beyond one year. Jenderny (2016) studies the top income mobility within 3 years with tax record data for Germany and finds that the top income mobility was fairly constant over the period 2001-2006 and persistence rates for top were somewhat higher than in Canada, the US and France. Evidence from Norway indicates that their top income mobility has increased somewhat in the 1990s (Aaberge et al., 2013). For the US and Canada, the concentration over time of income to the richest percent for the US and richest 0.01 percent for Canada has been stable, being around 60-70 percent (Saez and Veall, 2005; Auten et al., 2013).

The income mobility among the Finnish top is briefly discussed in Suoniemi

Figure 4.11 (Cont.) Fitting a Pareto distribution to the top of the income distribution: log-log plot and estimated alphas



Notes: The income concept is gross income excluding realized capital gains. In the left panel: vertical lines represent the top 1% threshold from joint, women's and men's distribution. The linear fit is estimated for top 5%. In the right panel: dotted line represents the income threshold (from the joint distribution) for top 10%, dashed line for top 5% and solid line for top 1%. The alpha parameters are estimated with maximum likelihood.

and Rantala (2010). The overall income mobility is shown to be significant but declining in the early 00s. In the top groups approximately 66 percent were the same individuals after 5 years but this decreased to 54 percent in the early 00s. However, the data used in their study is top-coded so they cannot study the mobility within the top 1% group properly. This section tackles the question of income mobility with an improved dataset for women and men separately.

The income mobility measure here is defined as the proportion of individuals who stay in a certain income group after 1, 5 or 10 years and is called persistence rate as in Jenderny (2016). This captures the movement downwards of the certain top income group. To avoid the sizes of the groups affecting the results, I also use equal group sizes within the top 10% and top 1%[20]. The proportion to stay in a certain group is conditional on being in the same income group on all periods being studied. The income groups are defined for both genders from the joint income distribution.

In figure 4.12 the persistence rates after one year are presented. The proportion of individuals above the highest percentile, 0.1%, is the most transitory and volatile. For men, this rate has ranged from 60 percent to almost 75 % while for women the very top membership is more volatile ranging from 50 % and 75 %. The other groups are more stable and exhibit a clear difference between the genders in all groups. The top 10% consists 83 % of same men after one year, while there are just below 80 % of same women in this group.

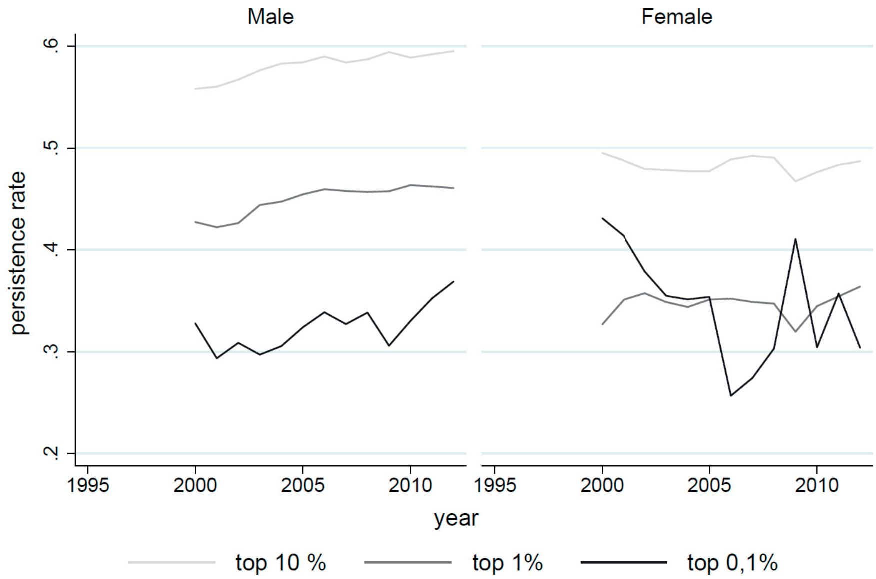
Figure 4.12 Persistence rates for top groups, after 1 year

Notes: Income measure is gross individual income excluding realized capital gains.

Figures 4.13 and 4.14 show the persistence rates for 5 and 10-year-periods over time. For these longer time periods there is a stark difference between

[20]Jenderny (2016) also discusses this point and proposes to use equal sizes or use rank statistics.

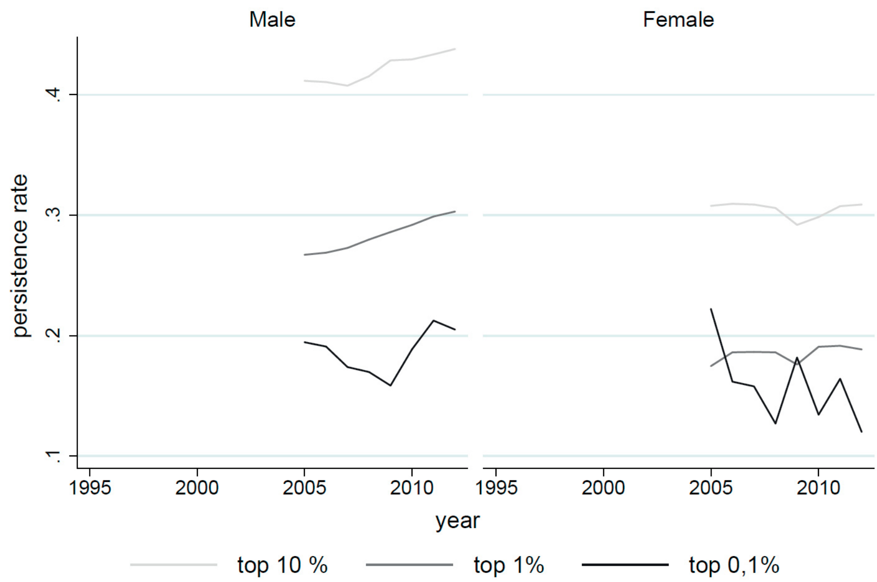
Figure 4.13 Persistence rates for top groups, 5 years



Notes: Income measure is gross individual income excluding realized capital gains. Conditional on being in the top group each year.

the genders and the picture from the one year mobility is somewhat different. For women, the top 1% and top 0.1% are on average similar but for men there is a large difference in the rates between these two groups. The difference between the other groups are larger than in the previous figure. In the top 10% approximately 50 % of the women stay in this top group after 5 years but for men this rate has increased from 56 % to 60 %. The persistence of men at the overall top over the 10-year horizon has increased somewhat and in the top 1% there were approximately 30 percent of same men each year since 2000. For women, there is no similar increasing trend in the persistence and the respective persistence rate is much lower, in approximately 19 %.

Figure 4.14 Persistence rates for top groups, 10 years



Notes: Income measure is gross individual income excluding realized capital gains. Conditional on being in the top group each year.

In the above figures, one shortcoming is that the group sizes vary which mechanically increases the downward mobility in the smaller groups. To see if richer people are genuinely prone to more downward mobility, I split the top groups to equal sizes and compare these persistences between genders. The results are presented with respect to only downward mobility, that is I calculate the individuals who has the same or higher rank in year t compared to t-1 or t-5. These results are presented in tables 4.4 and 4.5.

The persistence in equal size groups within top 10% reveal that higher incomes are less mobile. Also the gender divide is present, women are more likely to move downward. In the top 1% the gender differences are not so clear but the mobility also decreases in the upper tail. It is noticeable that since the equal size groups are formed from the joint distribution, the gender-specific groups are not equal sizes as women are under-represented in the top. However, from the overall analysis in this section we can conclude that there is sort of a "paper floor effect" (Guvenen et al., 2014) for women present in Finland which has not disappeared over time.

Table 4.4 Persistence within top 10%, after 1 year (left) and 5 years (right)

Persistence between t and t-1						
	1996		2005		2012	
Rank	Women	Men	Women	Men	Women	Men
1	.498	.5	.472	.518	.487	.534
2	.487	.496	.491	.509	.520	.529
3	.511	.523	.509	.543	.474	.550
4	.513	.511	.478	.513	.510	.544
5	.547	.530	.531	.542	.509	.549
6	.551	.534	.559	.566	.558	.566
7	.546	.547	.553	.579	.546	.581
8	.558	.579	.562	.605	.581	.609
9	.591	.656	.614	.647	.624	.654
10	.657	.731	.665	.745	.688	.764

Persistence between t and t-5						
	1996		2005		2012	
Rank	Women	Men	Women	Men	Women	Men
1	.410	.418	.332	.411	.313	.401
2	.414	.406	.349	.392	.306	.402
3	.423	.406	.338	.400	.291	.396
4	.414	.418	.334	.396	.285	.372
5	.406	.414	.323	.395	.297	.397
6	.412	.401	.350	.394	.315	.387
7	.393	.422	.338	.410	.319	.384
8	.373	.423	.331	.428	.317	.418
9	.340	.460	.354	.433	.345	.44
10	.418	.510	.453	.537	.474	.553

Table 4.5 Persistence within top 1%, after 1 year (left) and 5 years (right)

Persistence between t and t-1						
	1996		2005		2012	
Rank	Women	Men	Women	Men	Women	Men
1	.327	.332	.384	.424	.326	.451
2	.333	.405	.449	.421	.369	.409
3	.268	.385	.413	.442	.346	.456
4	.393	.403	.450	.448	.324	.437
5	.407	.425	.333	.418	.382	.436
6	.421	.437	.314	.4	.346	.399
7	.346	.469	.474	.423	.420	.453
8	.365	.478	.436	.480	.397	.495
9	.489	.577	.508	.546	.5	.492
10	.709	.676	.727	.648	.637	.716

Persistence between t and t-5						
	1996		2005		2012	
Rank	Women	Men	Women	Men	Women	Men
1	.25	.356	.243	.306	.304	.310
2	.303	.348	.159	.298	.202	.256
3	.338	.347	.266	.304	.204	.303
4	.287	.319	.225	.309	.220	.275
5	.291	.365	.253	.286	.259	.284
6	.169	.351	.259	.294	.217	.255
7	.211	.352	.305	.271	.202	.269
8	.354	.359	.236	.303	.176	.338
9	.333	.355	.254	.322	.193	.372
10	.370	.381	.424	.363	.337	.450

4.6 Discussion

This paper has analysed Finnish top incomes from a gender perspective. The analysis contributes to the literature by using a unique dataset without top-coding and extending the top income literature to the direction of gender issues. I discovered that the share of women in the top group (defined as the top 10%) is slightly less than 30 percent and there has been few changes in this share during the period 1995-2012. However, within this group women have got richer, which shows up as increasing representation in the top 1%. Comparing to other countries (Atkinson et al., 2018; Roine et al., 2017) it is interesting to remark that the gender divide at top is similar even though the institutions differ. In an international comparison Finland does not outperform the other countries even when in general Finland is considered as one of the most gender-equal countries.

I also discovered that there are clear gender-specific differences in the individual income composition, income distribution and income dynamics. In the top 1% and above, capital income is on average more important for women. However, in the most recent years, the share of wage income has increased at the top for women. Women also held more upper management positions at the end of the observation period compared to the 1990s. This indicates that the share of women who work their way upwards in the income ladder is on the rise. Nonetheless, women are also more likely to drop from the top income groups than men but the downward mobility decreases for both genders in higher income ranks. The gender-specific income distributions show that the income among women is less dispersed but the inequality has increased over time.

Several explanations for increased interest in top incomes can be offered. Top incomes are an important research field as high incomes relate to political and bargaining power that affects other parts of the income distribution as well and have global significance. With the help of improved data it is possible to explore the reasons behind growing disparities. From an overall inequality perspective, Roine and Waldenström (2015) show that top income shares are in fact associated positively and strongly with several inequality measures. The gender aspects of top incomes is of great importance because strong claims cannot be made about gender equality with respect to income without studying the whole distribution. Also, even in a country like Finland, gender equality is a topical question and thus needs to be monitored.

This paper, as a first step, has provided descriptive evidence on income and women. The next important question is to determine what lies behind it. There are several theories explaining the growing inequality especially at the upper end of the distribution (Alvaredo et al., 2013) but since the dynamics of the phenomenon is complex, the precise quantification of different causal factors is difficult. The gender dimension makes further analysis even more of a challenge because there are many unobservable factors affecting the positions of women and men in the labour markets and in the economy. For example, we do not know how the perception of success in the labour market differs between genders, a factor that might explain differences in incomes. Also, even when we have some knowledge about the gender norm effects on the labour market participation in the US (Bertrand et al., 2015), the Nordic countries have very different institutions.

The extent of assortative mating is an important question but with the current dataset the analysis of this theme was incomplete. Another potential mechanism in the evolution of female top incomes is the income shifting within households. Yet another fruitful avenue would be to analyse the dependencies between earnings and capital distributions, i.e. copulas theory. These questions should be studied further in the future.

Appendix

Major changes in Finnish top income taxation 1995-2012

Between years 1995-2012 the personal taxation has faced several reforms. The main trend in the tax reforms has been to reduce the tax rates and broaden the tax base. While the labour income taxation concerning the top income receivers has gone through a minor reforms over time, the capital and wealth income taxation has faced a more significant reforms.

From 1990 onwards until 2005, the corporation taxes were fully imputed (avoir fiscal system). The meaning here was to remove the double taxation of dividends and certain kind of interest incomes. In this system, the corporation tax base included the dividends and interests which were payed out to the owners. The individual receiving the dividend or interest payment could then reduce his own tax burden with the same amount that the corporation had paid. This meant that if the dividend was capital income (after 1993), the individual did not have to pay any tax for this income, as the two tax rates (capital and corporation) were at the same level.

In 1993 Finland started to apply the dual income tax system. The earned income (wages, benefits, pensions, transfers, earnings shares, i.e. items not listed as capital) is taxed at a progressive tax rate and capital income (interests, part of dividends and realized gains, rents, insurance income, enterprise capital share, forest capital share) at a flat tax rate. The tax rate on capital (and corporations) was 25 per cent in 1995, 28 per cent between years 1996-1999, 29 per cent between years 2000-2004 and 28 per cent between years 2005-2011. Since 2012 there has been two tax rates for capital income, first set to 30 per cent for income that was less than 50 000 € and 32 per cent for income that was over the threshold.

After avoir fiscal system was abolished, there was a shift towards the partial double taxation of dividends. Part of the dividends were tax-free under personal taxation. 70 per cent of the dividends from publicly listed companies were included in the personal capital income base and the rest was tax free. Dividends from privately held businesses are assigned as capital or labour incomes depending on the amount of dividend and net wealth of the business. If the return on the shares was less than 9 % of the firms net wealth and the dividends was below 90 000 € (60 000 € after 2011), the receiver paid no taxes. The dividends exceeding these thresholds were 30% tax-free and 70% taxable under capital income taxation. If the return on the shares were more than 9 % of the net worth, the exceeding amount was taxable under labour taxation for the 70% part and tax-free for 30%. Since 2005 the corporation tax rates and capital income tax rates have not moved hand in hand anymore, in fact the corporation tax rates are much lower.

The common interests are under tax-at-source since 1991. These are not part of the income statistics. Wealth tax was abolished from the beginning of 2006.

Income concepts

The incomes are drawn from the Statistics Finland's total statistics where the disposable income concept differs from the income distribution statistics. The primary difference is that the income concept used here includes the taxable realized capital gains.

Income items	Notes
gross labour income	
=regular wages and salary	
+ benefits in kind	
+ overtime compensation	
- pay generating costs	excl. travel expenses
+ gross self-employment income	
=income from agriculture	incl. property income
+ net forestry income	
+ other self-employment income	incl. property income
+ income from immaterial rights	
+ gross capital income	excl. interests taxed
=dividends income	
+ realized capital gains	
- realized losses	
= FACTOR INCOME	
+ income transfers	
=pensions	private and public
+ sickness benefits	
+ insurance payments	
+ unemployment benefits	
+ other transfers	
= GROSS INCOME	
- income taxes, social contributions	excl. tithe
- labour income taxes in municipal and state taxation	
- capital income taxes in state taxation	
- taxes from self-employment	
- other mandatory contribution	
- wealth tax	until 2005
= DISPOSABLE INCOME	

Additional tables

Table 4.A1 Descriptive statistics on income items

	Men			Women		
	mean	standard deviation	ind.	mean	standard deviation	ind.
1995						
Gross income	23 849.17	24 431.48	178 085	18 084.71	13 476.35	194 606
Disposable income	16 426.40	16 606.88	178 085	13 533.64	8 927.10	194 606
Wage income	14 485.12	18 323.17	178 085	9 876.44	12 290.99	194 606
Self-emp. income	2 330.47	9 755.95	178 085	824.15	6 162.23	194 606
Capital income	885.52	16 006.93	178 085	434.93	7 151.84	194 606
Realized capital gains (RCG)	322.15	12 438.97	178 085	134.82	4 737.44	194 606
Transfers	6 148.07	8 482.74	178 085	6 949.20	6 369.24	194 606
Gross income excl. RCG	23 527.03	19 822.56	178 085	17 949.89	12 337.22	194 606
1996						
Gross income	24 442.94	26 776.74	178 432	18 418.17	13 033.53	194 860
Disposable income	16 813.55	18 069.90	178 432	13 752.92	8 207.80	194 860
Wage income	15 137.96	18 938.34	178 432	10 356.27	12 777.38	194 860
Self-emp. Income	2 120.13	9 025.01	178 432	758.60	5 853.67	194 860
Capital income	1 009.93	19 519.87	178 432	447.00	5 818.30	194 860
Realized capital gains (RCG)	363.85	16 272.35	178 432	113.21	2 813.06	194 860
Transfers	6 174.91	8 688.28	178 43	6 856.31	6 399.90	194 860
Gross income excl. RCG	24 079.08	20 140.47	178 432	18 304.95	12 474.84	194 860
1997						
Gross income	25 373.32	27 035.91	179 433	18 831.36	14 757.12	195 508
Disposable income	17 753.63	18 834.32	179 433	14 227.19	9 711.26	195 508
Wage income	15 876.14	19 679.58	179 433	10 713.97	13 077.80	195 508
Self-emp. Income	2 344.89	10 378.41	179 433	816.22	7 169.64	195 508
Capital income	1 247.40	18 311.88	179 433	570.31	7 550.40	195 508
Realized capital gains (RCG)	473.01	12 856.40	179 433	194.71	3 578.34	195 508
Transfers	5 904.90	8 554.37	179 433	6 730.87	6 366.90	195 508
Gross income excl. RCG	24 900.31	22 612.70	179 433	18 636.65	13 972.74	195 508
1998						
Gross income	26 532.13	36 266.37	180 221	19 314.02	16 812.73	196 261
Disposable income	18 487.14	23 696.88	180 221	14 533.96	12 019.16	196 261
Wage income	16 870.06	25 051.32	180 221	11 104.89	13 470.49	196 261
Self-emp. Income	2 303.59	10 247.49	180 221	797.32	5 693.20	196 261
Capital income	1 590.96	25 964.59	180 221	714.68	11 450.76	196 261
Realized capital gains (RCG)	683.47	21 973.56	180 221	272.96	6 228.91	196 261
Transfers	5 767.52	8 793.93	180 221	6 697.15	6 650.08	196 261
Gross income excl. RCG	25 848.66	27 503.61	180 221	19 041.06	15 060.64	196 261
1999						
Gross income	28 316.34	84 430.28	180 964	20 027.44	33 567.54	197 197
Disposable income	19 698.52	46 370.54	180 964	15 114.56	23 416.11	197 197
Wage income	17 961.50	70 113.69	180 964	11 484.06	17 841.83	197 197
Self-emp. Income	2 331.71	11 666.58	180 964	808.73	5 918.34	197 197
Capital income	2 268.61	44 761.98	180 964	1 053.90	28 920.09	197 197
Realized capital gains (RCG)	1 176.74	40 157.79	180 964	534.44	24 956.79	197 197
Transfers	5 754.52	9 057.43	180 964	6 680.78	6 631.10	197 197
Gross income excl. RCG	27 139.59	71 674.95	180 964	19 493.01	19 961.04	197 197
2000						
Gross income	29 295.78	158 111.70	181 960	20 351.85	41 220.93	197 671
Disposable income	20 343.87	75 889.04	181 960	15 350.01	28 725.41	197 671
Wage income	18 687.67	148 103.80	181 960	11 779.28	19 846.49	197 671
Self-emp. Income	2 355.33	10 547.42	181 960	827.68	6 059.22	197 671
Capital income	2 653.74	47 983.75	181 960	1 230.53	36 363.74	197 671
Realized capital gains (RCG)	1 386.22	44 295.56	181 960	632.57	31 762.64	197 671
Transfers	5 599.04	8 813.98	181 960	6 514.37	6 568.32	197 671
Gross income excl. RCG	27 909.57	149 444.90	181 960	19 719.29	21 738.05	197 671

Table 4.A1: Descriptive statistics on income items, cont.

	Men			Women		
	mean	standard deviation	ind.	mean	standard deviation	ind.
2001						
Gross income	28 746.85	54 934.64	183 110	20 395.26	22 418.86	198 642
Disposable income	20 373.97	33 433.41	183 110	15 640.21	17 979.2	198 642
Wage income	18 778.18	39 896.15	183 110	12 173.30	14 692.82	198 642
Self-emp. Income	2 250.93	10 262.49	183 110	819.25	6 233.17	198 642
Capital income	2 058.83	35 507.53	183 110	866.43	17 711.16	198 642
Realized capital gains (RCG)	649.68	30 385.08	183 110	211.67	4 336.40	198 642
Transfers	5 658.97	8 863.02	183 110	6 536.26	6 713.41	198 642
Gross income excl. RCG	28 097.17	43 348.68	183 110	20 183.58	21 280.70	198 642
2002						
Gross income	29 123.12	61 302.91	184 446	20 857.98	22 539.17	199 683
Disposable income	20 850.89	36 141.55	184 446	16 079.59	16 764.12	199 683
Wage income	18 872.72	51 584.28	184 446	12 483.00	16 378.91	199 683
Self-emp. Income	2 284.91	10 823.16	184 446	817.43	6 226.86	199 683
Capital income	2 070.74	31 392.89	184 446	890.63	16536.21	199 683
Realized capital gains (RCG)	532.67	21 793.02	184 446	205.30	6 947.577	199 683
Transfers	5 894.75	9 280.47	184 446	6 666.92	6 914.176	199 683
Gross income excl. RCG	28 590.45	56 221.53	184 446	20 652.68	20 847.02	199 683
2003						
Gross income	29 439.35	38 088.16	185 621	21 431.54	35 630.47	200 724
Disposable income	21 282.85	28 059.99	185 621	16 596.92	31 713.02	200 724
Wage income	18 884.19	25 358.20	185 621	12 770.26	15 020.87	200 724
Self-emp. Income	2 231.93	10 833.19	185 621	821.89	6 347.59	200 724
Capital income	2 238.89	27 049.01	185 621	1 040.41	32 831.1	200 724
Realized capital gains (RCG)	519.06	13 011.86	185 621	240.77	4 649.47	200 724
Transfers	6 084.34	9 351.26	185 621	6 798.97	7 003.70	200 724
Gross income excl. RCG	28 920.28	34 042.34	185 621	21 190.77	34 333.1	200 724
2004						
Gross income	30 774.13	54 233.35	186 835	22 239.27	34 705.44	201 587
Disposable income	22 389.6	41 978.93	186 835	17 290.21	29 936.48	201 587
Wage income	19 518.40	26 617.47	186 835	13 267.07	16 012.78	201 587
Self-emp. Income	2 226.92	11 438.36	186 835	823.58	6 303.94	201 587
Capital income	2789.28	459 59.11	186 835	1 199.67	31 447.27	201 587
Realized capital gains (RCG)	742.00	29 561.99	186 835	285.94	7 729.71	201 587
Transfers	6 239.54	9 500.58	186 835	6 948.95	7 058.77	201 587
Gross income excl. RCG	30 032.13	42 331.76	186 835	21 953.33	32 143.61	201 587
2005						
Gross income	31 432.23	47 233.92	188 196	22 788.03	37 467.06	202 676
Disposable income	22 737.96	32 464.81	188 196	17 618.90	27 844.65	202 676
Wage income	20 104.56	28 194.18	188 196	13 678.13	16 365.57	202 676
Self-emp. Income	2 294.226	11 724.75	188 196	842.20	6 822.31	202 676
Capital income	2 734.68	36 493.14	188 196	1 230.217	34 309.48	202 676
Realized capital gains (RCG)	917.23	27 116.37	188 196	415.48	23 766.15	202 676
Transfers	6 298.76	9 598.30	188 196	7 037.47	7 169.50	202 676
Gross income excl. RCG	30 514.99	34 726.93	188 196	22 372.55	24 672.98	202 676
2006						
Gross income	32 602.45	77 706.96	189 162	23 283.96	46 166.02	203 564
Disposable income	23 595.48	53 827.63	189 162	18 059.09	33 142.97	203 564
Wage income	20 784.22	34 946.61	189 162	13 984.39	17 240.99	203 564
Self-emp. Income	2 279.43	11 889.13	189 162	868.86	7 793.89	203 564
Capital income	3 192.87	68 265.04	189 162	1 342.80	43 178.40	203 564
Realized capital gains (RCG)	1 422.26	63 819.02	189 162	562.42	39 132.76	203 564
Transfers	6 345.94	9 761.04	189 162	7 087.89	7 282.32	203 564
Gross income excl. RCG	31 180.19	39 445.43	189 162	22 721.54	20 021.14	203 564

Table 4.A1: Descriptive statistics on income items, cont.

	Men			Women		
	mean	standard deviation	ind.	mean	standard deviation	ind.
2007						
Gross income	33 878.47	61 372.77	190 225	23 741.41	22 234.53	204 479
Disposable income	24 694.12	41 259.53	190 225	18 513.12	15 218.33	204 479
Wage income	21 355.52	37 437.83	190 225	14 356.42	17 939.89	204 479
Self-emp. Income	2 576.14	13 033.17	190 225	944.13	7 503.72	204 479
Capital income	3 579.39	45 194.79	190 225	1 350.02	14 103.93	204 479
Realized capital gains (RCG)	1 618.48	35 647.42	190 225	538.25	8 910.60	204 479
Transfers	6 367.42	9 959.44	190 225	7 090.83	7 362.87	204 479
Gross income excl. RCG	32 260.00	45 148.83	190 225	23 203.16	19 059.69	204 479
2008						
Gross income	33 183.39	52 508.13	191 631	23 881.18	33 250.17	205 709
Disposable income	24 376.7	36 049.08	191 631	18 725.67	23 599.75	205 709
Wage income	21 529.11	32 992.06	191 631	14 728.93	17 579.40	205 709
Self-emp. Income	2 361	11 707.65	191 631	914.99	7 477.75	205 709
Capital income	2 885.797	39 229.37	191 631	1 179.053	28 961.92	205 709
Realized capital gains (RCG)	827.77	27 413.85	191 631	359.70	26 790.21	205 709
Transfers	6 407.48	10 025.93	191 631	7 058.20	7 399.89	205 709
Gross income excl. RCG	32 355.62	42 405.03	191 631	23 521.48	19 130.40	205 709
2009						
Gross income	32 538.35	41 686.72	192 781	24 402.18	22 242.49	206 678
Disposable income	24 315.83	28 567.63	192 781	19 357.87	15 633.54	206 678
Wage income	20 756.65	30 112.24	192 781	14 978.24	17 757.11	206 678
Self-emp. Income	2 039.986	10 468.38	192 781	843.40	7 118.38	206 678
Capital income	2 520.71	27 373.71	192 781	1 036.73	15 301.07	206 678
Realized capital gains (RCG)	571.55	15 534.11	192 781	259.41	11 424.75	206 678
Transfers	7 221.01	10 570.30	192 781	7 543.81	7 852.34	206 678
Gross income excl. RCG	31 966.80	36 678.86	192 781	24 142.76	18 263.93	206 678
2010						
Gross income	33 444.27	46 387.91	194 116	24 955.60	30 209.25	207 939
Disposable income	24 926.83	32 977.38	194 116	19 711.78	21 609.74	207 939
Wage income	20 748	27 045.78	194 116	15 121.22	18 551.04	207 939
Self-emp. Income	2 162.58	11 936.91	194 116	878.46	7 740.90	207 939
Capital income	3 088.66	35 973.95	194 116	1 267.61	24 545.77	207 939
Realized capital gains (RCG)	941.51	24 013.86	194 116	403.72	17 747.64	207 939
Transfers	7 445.03	10 798.48	194 116	7 688.31	7 933.07	207 939
Gross income excl. RCG	32 502.76	36 154.23	194 116	24 551.88	20 650.54	207 939
2011						
Gross income	33 948.14	104 285.10	195 295	24 971.76	27 059.80	208 923
Disposable income	25 275.19	75 195.20	195 295	19 711.07	19 235.51	208 923
Wage income	20 956.06	27 703.1	195 295	15 138.08	18 485.98	208 923
Self-emp. Income	2 196.028	12 254.84	195 295	909.79	7 869.95	208 923
Capital income	3 469.56	99 586.63	195 295	1 304.282	20 508.62	208 923
Realized capital gains (RCG)	1 189.019	88 146.57	195 295	388.45	14 324.62	208 923
Transfers	7 326.49	10 813.49	195 295	7 619.60	7 924.53	208 923
Gross income excl. RCG	32 759.12	41 015.78	195 295	24 583.30	21 803.60	208 923
2012						
Gross income	33 092.08	40 783.97	196 967	25 032.32	22 195.65	210 288
Disposable income	24 572.82	28 086.85	196 967	19 732.23	15 082.27	210 288
Wage income	20 850.40	26 704.07	196 967	15 225	18 634.98	210 288
Self-emp. Income	2 051.664	11 383.39	196 967	877.20	7 992.11	210 288
Capital income	2 617.73	29 318.28	196 967	1117.12	13 660.51	210 288
Realized capital gains (RCG)	522.94	13 324.35	196 967	240.75	4 027.85	210 288
Transfers	7 572.28	11 040.32	196 967	7 813.00	8 104.217	210 288
Gross income excl. RCG	32 569.14	36 110.50	196 967	24 791.57	21 495.90	210 288
Observations over time			3 357 480			3 626 995

Notes: income items deflated to 2008 euros.

Table 4.A2 Marginal effects from logistic regression, , dependent variable being in the top 10%, full list

VARIABLES	Women (single)	Men (single)	Women (cohabiting)	Men (cohabiting)	Women (cohabiting)	Men (cohabiting)
Working as professional in						
finance	0.0533***	0.0656***	0.0599***	0.122***	0.0629***	0.127***
	(0.00457)	(0.00814)	(0.00338)	(0.00793)	(0.00385)	(0.00827)
legal services	0.0187***	-0.00390	0.0225***	-0.0106***	0.0219***	-0.0100***
	(0.00274)	(0.00254)	(0.00217)	(0.00265)	(0.00243)	(0.00290)
health services	0.0166***	0.0189***	0.0122***	0.0368***	0.0146***	0.0391***
	(0.00247)	(0.00588)	(0.00182)	(0.00591)	(0.00213)	(0.00627)
Education level						
secondary level	0.0123***	0.0135***	0.00989***	0.0439***	0.00607***	0.0345***
	(0.00180)	(0.00297)	(0.00140)	(0.00317)	(0.00168)	(0.00368)
lowest level tertiary	0.0248***	0.0364***	0.0266***	0.0731***	0.0243***	0.0634***
	(0.00155)	(0.00254)	(0.00123)	(0.00256)	(0.00145)	(0.00287)
lower-degree level tertiary	0.0400***	0.0680***	0.0498***	0.140***	0.0466***	0.126***
	(0.00225)	(0.00341)	(0.00187)	(0.00333)	(0.00205)	(0.00358)
higher-degree level tertiary	0.115***	0.122***	0.141***	0.250***	0.137***	0.227***
	(0.00343)	(0.00454)	(0.00269)	(0.00430)	(0.00288)	(0.00461)
doctorate or equivalent	0.205***	0.192***	0.255***	0.386***	0.251***	0.359***
	(0.0122)	(0.0132)	(0.00858)	(0.00986)	(0.00866)	(0.0101)
education in STEM	0.0116***	0.0116***	0.0223***	0.0211***	0.0268***	0.0268***
	(0.00298)	(0.00299)	(0.00245)	(0.00316)	(0.00281)	(0.00341)
Other control variables (own)						
married	0.00295	0.0342***	0.00286**	0.0289***	-0.00217	0.0256***
	(0.00204)	(0.00280)	(0.00129)	(0.00203)	(0.00158)	(0.00229)
children	0.0194***	0.0265***	0.0116***	0.0375***	0.0107***	0.0354***
	(0.00142)	(0.00256)	(0.00114)	(0.00183)	(0.00134)	(0.00208)
small children	-0.0139***	-0.0171***	-0.0188***	0.00314*	-0.0216***	0.0106***
	(0.00212)	(0.00490)	(0.00113)	(0.00181)	(0.00130)	(0.00197)
living in the capital region	0.0323***	0.0303***	0.0386***	0.0752***	0.0350***	0.0708***
	(0.00152)	(0.00199)	(0.00143)	(0.00235)	(0.00161)	(0.00259)
age	0.00693***	0.0101***	0.0154***	0.0243***	0.0170***	0.0212***
	(0.000278)	(0.000379)	(0.000425)	(0.000503)	(0.000812)	(0.00106)
age squared	-4.69e-05***	-7.22e-05***	-0.000139***	-0.000203***	-0.000161***	-0.000181***
	(2.90e-06)	(4.29e-06)	(4.75e-06)	(5.42e-06)	(8.66e-06)	(1.12e-05)
native finnish	0.0314***	0.0527***	0.0449***	0.144***	0.0458***	0.143***
	(0.00303)	(0.00397)	(0.00249)	(0.00468)	(0.00341)	(0.00616)
native swedish	0.0404***	0.0609***	0.0419***	0.164***	0.0391***	0.149***
	(0.00416)	(0.00516)	(0.00325)	(0.00586)	(0.00456)	(0.00791)
self-employed	0.00206	-0.0201**	0.0382***	0.0130**	0.0405***	0.0294***
	(0.0159)	(0.00854)	(0.00625)	(0.00616)	(0.00698)	(0.00715)
upper management	0.187***	0.270***	0.165***	0.384***	0.172***	0.405***
	(0.0176)	(0.0115)	(0.00736)	(0.00706)	(0.00803)	(0.00799)
senior employees in R	0.0119	0.0379***	0.0142**	0.119***	0.0289***	0.152***
	(0.0160)	(0.00931)	(0.00617)	(0.00669)	(0.00696)	(0.00767)
senior employees in education	-0.0267*	-0.0618***	-0.0239***	-0.0673***	-0.00872	-0.0364***
	(0.0154)	(0.00917)	(0.00565)	(0.00673)	(0.00646)	(0.00777)
other senior employees	0.0448***	0.0579***	0.0537***	0.172***	0.0696***	0.199***
	(0.0158)	(0.00997)	(0.00611)	(0.00741)	(0.00690)	(0.00831)
supervisors	-0.0167	-0.00170	-0.00578	0.0217***	0.0114	0.0503***
	(0.0160)	(0.00887)	(0.00613)	(0.00632)	(0.00703)	(0.00740)
clerical workers, independent	-0.0647***	-0.0530***	-0.0503***	0.0252***	-0.0345***	0.0526***
	(0.0152)	(0.00851)	(0.00549)	(0.00649)	(0.00631)	(0.00751)
clerical workers, routine	-0.0917***	-0.112***	-0.0628***	-0.0794***	-0.0503***	-0.0528***
	(0.0155)	(0.0103)	(0.00591)	(0.0111)	(0.00683)	(0.0122)
lower-level admin. & clerical occ.	-0.0930***	-0.0797***	-0.0792***	-0.0930***	-0.0682***	-0.0672***
	(0.0152)	(0.00833)	(0.00542)	(0.00612)	(0.00623)	(0.00716)
workers in agriculture	-0.101***	-0.150***	-0.0914***	-0.165***	-0.0801***	-0.132***
	(0.0171)	(0.00878)	(0.00614)	(0.00992)	(0.00761)	(0.0114)
manufacturing workers	-0.0647***	-0.0572***	-0.0562***	-0.0559***	-0.0375***	-0.0245***
	(0.0165)	(0.00808)	(0.00631)	(0.00597)	(0.00758)	(0.00714)
other production workers	-0.101***	-0.121***	-0.0851***	-0.146***	-0.0732***	-0.117***
	(0.0155)	(0.00828)	(0.00577)	(0.00649)	(0.00675)	(0.00770)
distribution and service workers	-0.107***	-0.132***	-0.0837***	-0.162***	-0.0698***	-0.134***
	(0.0152)	(0.00800)	(0.00549)	(0.00621)	(0.00637)	(0.00740)
students	-0.115***	-0.171***	-0.0863***	-0.223***	-0.0755***	-0.200***
	(0.0152)	(0.00770)	(0.00559)	(0.00625)	(0.00644)	(0.00736)
pensioners	-0.108***	-0.155***	-0.0797***	-0.169***	-0.0676***	-0.136***
	(0.0152)	(0.00762)	(0.00542)	(0.00563)	(0.00626)	(0.00677)
unemployed	-0.115***	-0.170***	-0.0913***	-0.214***	-0.0799***	-0.186***
	(0.0151)	(0.00757)	(0.00540)	(0.00554)	(0.00621)	(0.00665)
unknown	-0.0925***	-0.146***	-0.0350***	-0.102***	-0.0195**	-0.0685***
	(0.0154)	(0.00790)	(0.00668)	(0.00732)	(0.00759)	(0.00852)

Table 4.A2: Marginal effects from logistic regression, dependent variable being in the top 10%, full list (cont.)

VARIABLES	Women (single)	Men (single)	Women (cohabiting)	Men (cohabiting)	Women (cohabiting)	Men (cohabiting)
spouse characteristics						
self-employed					0.00418	-0.0145*
					(0.00447)	(0.00762)
upper management					-0.0243***	-0.0380***
					(0.00443)	(0.00812)
senior employees in R& D					-0.0230***	-0.0431***
					(0.00441)	(0.00769)
senior employees in education					-0.0210***	-0.0480***
					(0.00467)	(0.00753)
other senior employees					-0.0196***	-0.0391***
					(0.00448)	(0.00753)
supervisors					-0.0250***	-0.0284***
					(0.00451)	(0.00797)
clerical workers, independent					-0.0132***	-0.0261***
					(0.00457)	(0.00719)
clerical workers, routine					-0.0130*	-0.00932
					(0.00734)	(0.00845)
lower-level admin. and clerical occ.					-0.0204***	-0.0423***
					(0.00464)	(0.00710)
workers in agriculture					-0.0269***	-0.0638***
					(0.00936)	(0.0140)
manufacturing workers					-0.0297***	-0.0350***
					(0.00451)	(0.00840)
other production workers					-0.0269***	-0.0502***
					(0.00521)	(0.00860)
distribution and service workers					-0.0231***	-0.0478***
					(0.00490)	(0.00765)
students					-0.0127**	-0.0506***
					(0.00546)	(0.00765)
pensioners					-0.00570	-0.0140*
					(0.00462)	(0.00771)
unemployed					-0.0133***	-0.0422***
					(0.00459)	(0.00733)
unknown					-0.0254***	-0.0562***
					(0.00539)	(0.00837)
age					-0.00171**	0.00361***
					(0.000730)	(0.00106)
age squared					2.24e-05***	-2.94e-05**
					(7.43e-06)	(1.16e-05)
native finnish					-0.00147	0.0139*
					(0.00521)	(0.00727)
native swedish					-0.000433	0.0370***
					(0.00604)	(0.00906)
education in STEM					-0.00653***	-0.00297
					(0.00186)	(0.00507)
secondary level					0.00538**	0.0134***
					(0.00256)	(0.00347)
lowest level tertiary					0.00588***	0.0261***
					(0.00189)	(0.00252)
lower-degree level tertiary					0.00308	0.0110***
					(0.00206)	(0.00299)
higher-degree level tertiary					0.000194	0.00766**
					(0.00204)	(0.00345)
doctorate or equivalent					-0.00784**	0.00375
					(0.00321)	(0.00828)

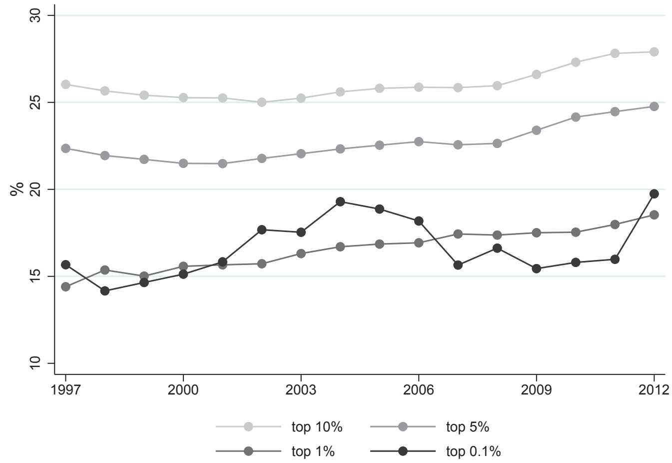
Table 4.A2: Marginal effects from logistic regression, dependent variable being in the top 10%, full list (cont.)

VARIABLES	Women (single)	Men (single)	Women (cohabiting)	Men (cohabiting)	Women (cohabiting)	Men (cohabiting)
spouse income group						
2nd decile					-0.0133***	-0.0317***
					(0.00270)	(0.00326)
3rd decile					-0.0150***	-0.0348***
					(0.00269)	(0.00333)
4th decile					-0.0115***	-0.0334***
					(0.00268)	(0.00336)
5th decile					-0.0130***	-0.0307***
					(0.00268)	(0.00343)
6th decile					-0.0101***	-0.0247***
					(0.00266)	(0.00347)
7th decile					-0.00538**	-0.0201***
					(0.00263)	(0.00355)
8th decile					0.00103	-0.00445
					(0.00260)	(0.00369)
9th decile					0.0186***	0.0186***
					(0.00265)	(0.00394)
10th decile					0.0392***	0.0517***
					(0.00280)	(0.00453)
top 1%					0.00534**	0.0538***
					(0.00269)	(0.0108)
Observations	1,017,288	964,654	1,017,288	964,654	812,841	810,482
Sample mean prob. to be in top 10%	0.0414	0.1159	0.0760	0.2448	0.0860	0.2656

Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Note: All predictors at their mean values. Year fixed effects also included in the regression.

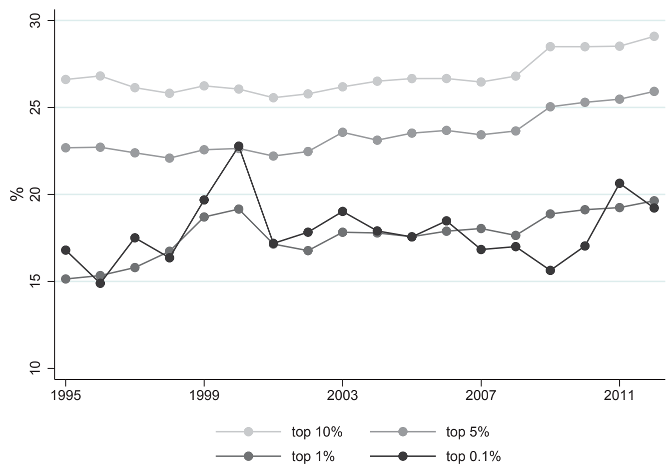
Additional figures

Figure 4.A1 Share of women in different income groups, 1995-2012



Notes: income distribution based on 3-year average gross income excluding the realized capital gains.

Figure 4.A2 Share of women in different income groups, years 1995-2012



Notes: income distribution based on gross income including the realized capital gains.

Bibliography

Aaberge, R., Atkinson, A., and Modalsli, J. (2013). The Ins and Outs of Top Income Mobility. *IZA Discussion paper No. 7729.*

Albrecht, J., Björklund, A., and Vroman, S. (2003). Is There a Glass Ceiling in Sweden? *Journal of Labor Economics*, 21(1):145–177.

Alstadsæter, A., Jacob, M., Kopczuk, W., and Telle, K. (2016). Accounting for Business Income in Measuring Top Income Shares: Integrated Accrual Approach Using Individual and Firm Data from Norway. Working Paper 22888, National Bureau of Economic Research.

Alvaredo, F., Atkinson, A., Piketty, T., and Saez, E. (2013). The Top 1 Percent in International and Historical Perspective. *Journal of Economic Perspectives*, 27:1–21.

Armour, P., Burkhauser, R. V., and Larrimore, J. (2013). Deconstructing Income and Income Inequality Measures: A Crosswalk from Market Income to Comprehensive Income. *American Economic Review*, 103(3):173–77.

Atkinson, A. B., Casarico, A., and Voitchovsky, S. (2018). Top incomes and the gender divide. *Journal of Economic Inequality*, 16:225–256.

Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1):3–71.

Auten, G., Gee, G., and Turner, N. (2013). Income Inequality, Mobility, and Turnover at the Top in the US, 1987-2010. *American Economic Review*, 103(3):168–72.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics*, 2(3):228–55.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender Identity and Relative Income within Households . *The Quarterly Journal of Economics*, 130(2):571–614.

Burkhauser, R. V., Hahn, M. H., and Wilkins, R. (2015). Measuring Top Incomes Using Tax Record Data: a Cautionary Tale from Australia. *Journal of Economic Inequality*, 13:181–205.

Cirillo, P. (2013). Are Your Data Really Pareto Distributed? *Physica A: Statistical Mechanics and its Applications*, 392(23):5947 – 5962.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703.

Edlund, L. and Kopczuk, W. (2009). Women, Wealth, and Mobility. *American Economic Review*, 99(1):146–78.

Fortin, N. M., Bell, B., and Böhm, M. (2017). Top Earnings Inequality and the Gender Pay Gap: Canada, Sweden, and the United Kingdom. *Labour Economics*, 47(Supplement C):107 – 123. EALE Conference Issue 2016.

Guvenen, F., Kaplan, G., and Song, J. (2014). The Glass Ceiling and The Paper Floor: Gender Differences among Top Earners , 1981 – 2012. NBER Working Paper No. 20560, The National Bureau of Economic Research.

Jenderny, K. (2016). Mobility of Top Incomes in Germany. *Review of Income and Wealth*, 62(2):245–265.

Jenkins, S. P. (2017). Pareto Models, Top Incomes and Recent Trends in UK Income Inequality. *Economica*, 84(334):261–289.

Jäntti, M., Riihelä, M., Sullström, R., and Tuomala, M. (2010). Trends in Top Income Shares in Finland. In Atkinson, A. B. and Piketty, T., editors, *Top Incomes, A Global Perspective*, chapter 8, pages 371–447. Oxford University Press.

OECD (2017). *The Pursuit of Gender Equality*. OECD Publishing.

Pietiläinen, M., editor (2013). *Työ, talous ja tasa-arvo*. Tilastokeskus, Statistic Finland.

Piketty, T. (2003). Income Inequality in France, 1901–1998. *Journal of Political Economy*, 111(5):1004–1042.

Piketty, T. and Saez, E. (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics*, 118(1):1–41.

Roine, J., Boschini, A., and Gunnarsson, K. (2017). Women in Top Incomes: Evidence from Sweden 1974 – 2013. Iza Discussion Paper Series No. 10979.

Roine, J. and Waldenström, D. (2015). Long-Run Trends in the Distribution of Income and Wealth. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution, volume 2A*, chapter 7, page 469–592. Elsevier B.V.

Saez, E. and Veall, M. (2005). The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence. *American Economic Review*, 95:831–49.

Selin, H. and Pirttilä, J. (2011). Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993. *The Scandinavian Journal of Economics*, 113(1):120–144.

Statistics Finland. Suomen virallinen tilasto (svt) (finnish official statistics): Tulonjaon kokonaistilasto (total statistics on income distribution), tuloerot 2014, laatuseloste: Tulonjaon kokonaistilasto. `http://www.stat.fi/til/tjkt/2014/02/tjkt_2014_02_2015-12-18_laa_001_fi.html`. Accessed: 2.1.2017.

Suoniemi, I. and Rantala, J. (2010). Income Mobility, Persistent Inequality and Age, Recent Experiences from Finland. *Labour Institute for Economic Research*, Discussion papers(263).

5 THE EFFECTS OF WORKING HOURS REDUCTION ON HEALTH AND LABOUR MARKET EXITS: EVIDENCE FROM THE FINNISH PART-TIME PENSION PROGRAM

Terhi Ravaska, *Labour Institute for Economic Research, Helsinki, Finland*

Abstract

In this essay[1] *I examine the effects of reduced working hours on health-related factors such as purchase of prescription drugs, sickness absence days and labour market exits of elderly workers. In assessing the causal relationships, I exploit the eligibility ages and the changes in these age limits in a part-time retirement scheme and utilize a difference-in-differences method and a fixed effects instrumental variable estimation.*

The eligibility age of part-time retirement was lowered by 2 years to 56 in 1998. In the difference-in-differences setting the treatment group are those eligible for part-time retirement in the age of 56 while the control group had eligibility age at the age of 58. I find that the mental illness drug purchases increased in the treatment group while for other outcomes the results are imprecise. In the instrumental variable setting I find that the take-up of part-time pension decreased the probability to purchase any drugs by 2.8 percentage points for the group of compliers while there is no evidence of a relevant effect on the probability to buy mental illness drugs. The reduction in the

[1] I would like to thank Jukka Pirttilä, Jutta Viinikainen, Maarit Olkkola, Tuuli Paukkeri, Luke Munford and Tuomas Kosonen for valuable comments as well as seminar participants in the Labour Institute for Economic Research and Allecon seminars in 2017 and the discussants in the Summer Meeting of Finnish Economist in Jyväskylä, IIPF conference in Tampere 2018 and EuHea meeting in Catania in 2018. I also thank Kela for partly financing this project.

risk of early labour market exits were substantial, on average 5 percentage points, in the group of compliers.

Keywords: Part-time pension, health, FE-IV, difference-in-differences
JEL classification: J26, I10

5.1 Introduction

Ageing nations are struggling to find ways to increase labour force participation. One of the policy responses has been to increase the retirement age but the health and work ability of the elderly have been a concern. Beside the disability pension expenditure there are public expenses stemming from sickness days, drug prescriptions and public health care and there are indirect costs in the form of loss of output. Beside the adjustments in the retirement ages the policy responses which aim to improve the health of the elderly workers are also vital. One response has been to promote flexible work conditions and gradual retirement schemes, such as part-time pension schemes.

A part-time pension program aims to prolong working careers by reducing work loads which in turn can improve health and work motivation. If it is effective policy, this should reduce the number of disability pension spells and sickness absences. It has also been proposed that decreasing the work burden at the end of the career could ameliorate the well-being in the full retirement. However, there is little knowledge of the actual effects of part-time work on health and well-being.

In this study I explore whether the part-time pension program and working hours reduction affects the health outcomes of elderly workers and reduces the risk of early labour market exits. The Finnish part-time pension scheme is suitable to study these questions because it certainly reduced the hours worked and was very generous so it had very modest effects on the disposable income or the future pension rights[2]. Also most of the part-time pensioners continued in their career job, so the effects are not driven by changes in the work community. The focus is on the short-term effects using register

[2]The scheme under study was abolished in the beginning of 2017 and replaced with flexible partial old-age pension with permanent reduction in pension but without requirement of reducing working hours.

data on health and labour market outcomes for the years 1995-2004.

I study these question in two empirical settings. Firstly, I exploit the changes in the eligibility age for the part-time pension. My control group are individuals who could retire at the age of 58 while in the treatment group the individuals were able to retire from the age of 56 onwards. In a difference-in-differences (DD) setting I am able to answer how the access (that is the intention-to-treat effects) to part-time pension earlier affected different health and labour market outcomes. I also study the take-up effect of part-time pension. As the part-time retirement decision is potentially endogenous, I instrument the decisions with the eligibility ages and account for the individual unobserved effects. This individual fixed effects instrumental variable (FE-IV) method reveals the effects for subgroup of compliers.

The data include all part-time pensioners born between 1940-1947 and who took part-time pension between 1998-2005. The primary health outcomes of interest are sickness days and the purchase of prescription drugs. The drug purchases and diagnosis data are on a monthly level and include the code for the drug or disease. The different causes of illness are taken into account by dividing the drug purchase data to represent the different aspects of mental and physical health. The sickness day spells data cover sickness absences of over 10 days. The dataset also includes administrative records on individual demographics and data on other pension spells as well as earnings information. I also study the effects of reduced hours on labour market exits via early retirement schemes.

I find that the drug purchases and the sickness absence spells respond to the part-time pension program or hours reduction. The DD estimates are imprecise but suggest that there was an increase in the drugs purchased due to the reform. For the take-up effect of the part-time pension I find that reducing working hours decreases the probability to purchase any drugs by 2.8 percentage points in a short term within a sub-group of compliers. This group has the worse health outcomes during the pre-pension period. In the number of packages purchased the average effect is a reduction of a quarter of a package within a year. The probability of long sickness absence one year after part-time retirement is reduced by 6.9 percentage points. The effects are stronger for women. The risk of early labour market exits is reduced by 5.5

percentage points.

This paper contributes to the literature on the relationship between work (or non-work) and health especially in the context of elderly workers. The previous literature has focused on full retirement and there is little knowledge of the issue at the intensive margin of working hours. The paper also contributes by evaluating the Finnish part-time pension scheme thoroughly with rich administrative data and with a suitable research design to estimate the causal effects. The set goals for the part-time pension were to improve health and work-ability and to reduce the risk of early labour market exits. However, the empirical evidence is scarce.

The paper proceeds as follows: section 5.2 summarizes the relevant literature. Section 5.3 goes through the institutional setting and the data used. Section 5.4 introduces the empirical strategies. Section 5.5 presents the results. Section 5.6 concludes.

5.2 Previous literature

There is some evidence that long working hours are positively associated with cardiovascular disease and stroke (Virtanen et al., 2012; Kivimäki et al., 2015), worse life habits measured as physical activity, smoking, alcohol consumption and diet (Taris et al., 2011) and depressive symptoms (Virtanen et al., 2018). For causal estimates, Cygan-Rehm and Wunder (2018) instrument the weekly working hours with statutory work-week hour regulation and show that one hour increase has adverse effects on health while Ahn (2016) show that reduced work week induce healthier life-style. Yet it is plausible that work hours have a different effect on health in the interval of part-time and normal work hours. The effects of reducing work hours from normal hours are a less studied topic.

The previous literature on the health effects of working on elderly workers mainly concern full retirement. However, there is at least one working paper explicitly studying a part-time work scheme in the late career on health (Kantarci and Kolodziej, 2017). In this study the finding is that part-time work in late career has positive effects on health compared to full-time work. Results from Dave et al. (2008) indicate that part-time work has a less negative effect on the health outcomes relative to the full-time retirement. The outcomes of these studies were survey answers so I contribute to this scarce literature by providing also evidence based on outcomes measured with register data.

The most common identification strategy in studies of retiring and health is based on instrumental variables where the rules of the retirement eligibility ages act as instruments. The identifying assumption is that health is affected by the instrument only indirectly through the effect of the actual retirement. The instrumental variable approach suggests that full-retirement has a negative effect on cognitive functioning (Rohwedder and Willis, 2010; Bonsang et al., 2012)[3], while Behncke (2012) finds negative health effects on general health. The identifying assumption in the instrumental variable approach is somewhat strong. Without a change in the age limit the individual can adjust her behaviour already before the actual age is reached. That means that the estimated effect is not a retirement effect *per se*. In the Finnish context there is also a sudden change in the part-time retirement age which mitigates this problem.

Another method recently used to reveal the causal effects between nonwork and health has been to utilize reforms in the pension schemes. Hallberg et al. (2015) study the targeted early retirement offer in Sweden which enabled some of the workers to retire at the age of 55 while others retired at the age of 60. Their conclusion is that early retirement leads to a reduced probability of dying conditional on age and reduced time spent in inpatient care. People with low pre-retirement incomes or low education benefit the most from the reduction in eligibility age. However, the reduction in the eligibility age in Norway had no effect on mortality (Hernaes et al., 2013). The mixed results might be caused by the fact that the Swedish reform was targeted to a specific group (the military personnel) while the Norwegian reform affected a larger part of the population. Hagen (2018) studies retirement effects on health among local government workers in Sweden. The reform he exploits caused a two-year increase in the normal retirement age. He finds no retirement effects on prescription drug purchases, mortality or hospitalizations.

[3] Rohwedder and Willis (2010) use cross-country data from several countries which limits the interpretability of the results because the cognitive functioning and institutions might be correlated and other unobserved differences between countries might influence the results.

5.3 Institutional framework and data

5.3.1 Finnish pension system

Part-time pension program

The part-time pension program was introduced in 1987 for private sector workers to increase the flexibility at the end of their careers. The age limit was set to 60 while old-age retirement age was 65. The public sector workers have been eligible for part-time pension since 1989 but their age limit was originally set to 58. In 1994 the eligibility ages were harmonized for both sectors and set to 58. In July 1998 the age limit was temporary reduced to 56 until the end of 2002 when the age was set back to 58.

The government lowered the age limit in 1998 in order to support the labour force participation of the elderly. The target was to reduce labour market exits through other pension programs. However, the reform was made temporary in order to receive information on whether this policy was effective. The reform in 1998 also added a clause which obliged employers to arrange part-time work whenever possible. (Government Bill HE 13/1998). However, employers were not monitored or sanctioned if part-time work was not made available.

Workers who fulfilled the working, pension accrual and age condition were eligible for part-time pension if part-time work was available. For a private sector worker the working condition required 12 months of full-time work during the 18 months preceding the part-time pension take-up and for the public sector the requirement was 6 months of full-time work during the preceding 18 months. The worker also needed to have been accruing pension rights for 5 years during the preceding 15 years in the private sector and 3 years during the preceding 5 years for the public sector. Also the employer needed to consent to the part-time work arrangement.

The pension provider monitored the income received during the part-time pension spell. The hours worked and the earnings needed to decrease in the same proportion being in the range of 30-75 percent from the previous levels. Work hours had to be at least 16 but at most 28 hours per week. The decrease in working time is not observable from the data. However, comparing the wage levels before and after part-time pension it seems that the reduction in hours is about 45 percent. This is also in line with the results from surveys done to the part-time pensioners (Takala, 2004).

The pension received during the part-time retirement was determined as half of the difference in earnings between full-time work and part-time work, however it couldn't exceed 75 percent of the accrued pension. The reduction in disposable income was not proportional to the decrease in earnings because the average tax rate for part-time pensioners was lower than for full-time pensioners or full-time workers conditional on income[4]. The difference also accrued future pension rights, with an accrual rate of 1,5 %. In full time work (and for the wages earned as part-time pensioner) accrual rates were 1,5% for the individuals below the age of 60 and 2,5% for the individuals who were 60 or older. All in all, the effect of part-time pension on life-time earnings was modest.

The part-time pension system has been evaluated in Kyyrä (2015). As in this study, he also exploits the changes in the eligibility ages. His findings suggest that part-time pension did not lengthen the work career and if anything slightly reduced the risk of unemployment or exit out of the labour market. However, these results hinge on the eligibility effect and not on the actual take-up of part-time pension. The current study can also look at the ones who have taken the part-time pension.

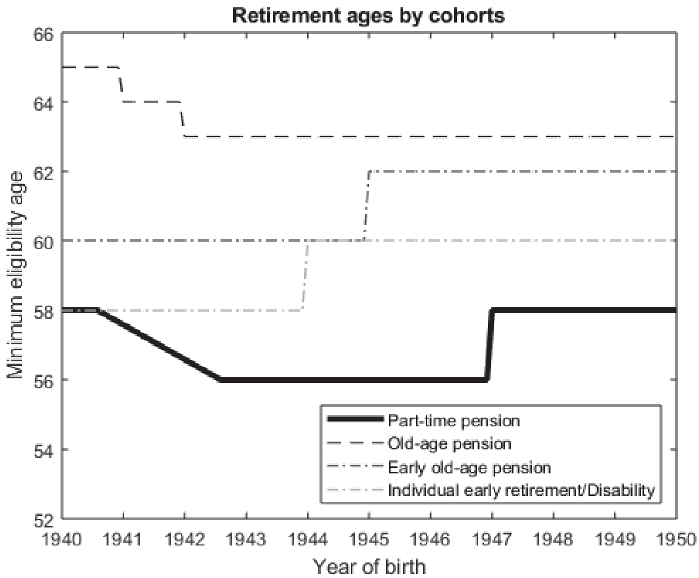
Other pension programs

The pension system at the end of the 1990s and early 2000s included multiple labour market exit pathways. Before the early old-age retirement age was achieved, individuals exited through disability, an individual early retirement scheme or unemployment pension. Figure 5.1 presents the eligibility ages by cohort for the different pension schemes[5]. For the cohort born in 1940 the eligibility age for individual early retirement had the same age limit as part-time pension. Individual early retirement was granted for individuals

[4]From the data, the ratio of net incomes between one year before the start of the part-time pension spell and one year after the starting year, conditional on being in the part-time pension for the whole year, is 88%.

[5]As part-time pensioners had to fulfil the work condition, the changes in the unemployment tunnel or unemployment pension are not shown or discussed.

Figure 5.1 Minimum eligibility ages for different pension schemes.



with reduced work ability but the medical conditions were more lenient than in the disability pension, which can be applied at any age. The pension reform in 2005 abolished the individual early retirement pension. However, it was also legislated that individuals over the age of 60 were to be granted disability pension with similar conditions as previously in the individual early retirement.

The individual early retirement scheme or the disability pension and the part-time pension were most likely partial substitutes. Individuals with strong preferences for working but limited work ability probably seek part-time pension possibilities. This fact already makes the part-time pensioners a selected sample from the whole population of elderly workers. However, not all individuals were entitled part-time work by their employers. This is another source of selection bias.

The major pension reform in the year 2005 affected the cohorts under study. The 2005 reform decreased the full old-age retirement age from 65 to 63. This reform might have had independent effects on health outcomes and it certainly affected labour market exit decisions. For this reason I focus on the period of 1995-2004 in this study.

5.3.2 Data sources

The data include various kinds of information collected from administrative registers governed by the Social Insurance Institution of Finland, Statistics Finland and Finnish Centre for Pensions for the years 1995-2014. The most important variables are drug purchases, including the code of disease, sickness benefit spells, pension benefit spells, educational and demographic characteristics as well as different income sources. The income measures are collected from the Finnish Centre for Pensions database and Statistic Finland's income data. The industry classification is based on the nationally modified version of the European Union's standard industrial classification (NACE classification). The education information includes the level of education in one-digit level.

The data include the exact date when the part-time pension spell has started or ended. The main sample includes all part-time pensioners for the years 1998-2005 who were born between 1940-1947. There are approximately 52 000 individuals in the part-time pensioners sample. Approximately 95 percent of the part-time pensioners have only one part-time pension spell and the average part-time pension spell is 4,4 years[6].

The drug purchases and the sickness benefit spells are retrieved from the registers of the Finnish Social Insurance Institution. The information about the drugs come from the purchase of prescription drugs -register which contains all the information of the prescription drugs sold by Finnish pharmacies. The information included is the drug code and day of purchase and covers the whole observation period. The sickness benefit -register includes the start and end day of the sickness benefit spell. Only sickness absences of over 10 days are observed in the data because the first 10 days are not covered by social insurance. The sickness absence benefit is earnings-related. For individuals who are on part-time retirement, the sickness absence benefit is calculated based on the earnings from the part-time work. On sickness leave they can continue taking the part-time pension.

The medicine data are classified with ATC (Anatomical Therapeutic Chem-

[6]Figure 5.A1 in the appendix shows the histogram of the length of the part-time pension spells.

ical) -codes. The outcome variables with respect to the drug data are the purchases of any positive amount of medicine (referred to as *extensive margin* in the result section) and the intensity of medicine purchased measured as the number of packages purchased during a year (referred to as *intensive margin* in the result section). The dataset is limited to observe the number of purchases but not the defined daily dose (DDD) which would more accurately describe the intensity of a disease. Neither is the size of the package observed so if drug companies have changed the sizes of the packages during the observation period this causes biases in the estimated effect at the intensive margin. During the observation period, the remuneration system changed very little.

Work amount can be related to health via several channels. For example, work can contribute to allostatic load which accumulates as an individual is exposed to chronic stress. On the other hand the increase of professional work has raised the question of how office work affects the individual's musculoskeletal system. For this reason I divide the medical data in subgroups representing different cause-specific sickness groups. Table 5.A1 in the appendix lists the classification used.

The part-time pensioners are not a representative subgroup of the population of elderly workers. Tables 5.A2 and 5.A3 in the appendix compare the part-time pensioners to a representative sample of workers from the same cohort and gender distribution ($\sim$ 150 000 individuals). On average, the part-time pensioners are more educated and earn more than their peers from the same cohorts. They are overrepresented in the fields of professional services, administration jobs and education. The baseline health outcomes are also more favourable for the part-time pensioners as they have fewer sickness absences and they purchase less medicine within a year.

5.3.3 Descriptives

Tables 5.1 and 5.2 show descriptive statistics for the treatment and the control groups which are used in the difference-in-differences regression models. The treatment group is those born between July 1942 and 1946 (eligibility age at 56) and the control group those born between 1940 and July 1940 (eligibility age at 58). These statistics concern the years 1995-1997 when nobody in the sample had taken the part-time pension yet and are measured when the individuals were aged 54-55. There are statistically significant differences in the income variables and some industries or professions between the two groups. These background variables will be controlled in the regression model.

The main health-related variables are sickness absences and any drug purchases or mental illness drug purchases. The treatment group has somewhat more sickness absences days before the part-time retirement period. Any drug purchases are similar between the groups, over half of the individuals have used some prescription medication at the ages of 54-55. These descriptives reveal that the two groups are quite similar but the parallel trends assumption is further studied in the next sections.

Table 5.1 Descriptives: Pre-reform and pre-part-time pension spell, years 1995-1997, treatment and control groups.

	Control	Treatment
Income and employment		
Wage income***,€	24 816.6 (11 600.5)	26 790.0 (15 521.3)
Net income***,€	17 856.6 (7 615.9)	20 469.3 (47 504.8)
Pension income,€	278.7 (1 560.5)	275.4 (1 559.8)
Health indicators		
Sickness absence days*	3.57 (15.28)	4.34 (18.74)
Any drug	0.58 (0.49)	0.57 (0.49)
Nr of purchases, any drugs	2.94 (4.45)	3.09 (4.86)
Drugs for respiratory diseases	0.25 (0.43)	0.25 (0.43)
Nr of purchases, resp. dis.	0.59 (1.67)	0.62 (1.88)
Drugs for circulatory diseases	0.21 (0.41)	0.23 (0.42)
Nr of purchases, circ. diseases*	1.00 (2.40)	1.13 (2.65)
Heart disease drugs	0.21 (0.41)	0.22 (0.42)
Nr of purchases, heart conditions*	0.99 (2.40)	1.12 (2.64)
Drugs for cerebro-vascular diseases	0.013 (0.12)	0.01 (0.11)
Nr of purchases, cerebrovas. diseases	0.03 (0.25)	0.03 (0.31)
Drugs for musculo-skeletal diseases	0.27 (0.44)	0.27 (0.44)
Nr of purchases, musculo-skeletal dis.	0.54 (1.206)	0.54 (1.20)
Drugs for diabetes	0.017 (0.13)	0.02 (0.14)
Nr of purchases, diabetes	0.08 (0.78)	0.10 (0.85)
Mental illness drugs*	0.12 (0.33)	0.11 (0.31)
Nr of purchases, mental disease	0.38 (1.40)	0.39 (1.90)
Individuals	2 880	9 508

Means with standard deviations in parentheses. All variables are measured within a year. Unit of measure for *drug purchases* is share of individuals with any purchases within a group and the *number of purchases* is measured in packages. *Sickness absence days* represent the absence days over 10 days. T-test for the difference of means, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The instrumental variable analysis estimates the effect for the compliers. For this reason I also present descriptive statistics for the group of compli-

Table 5.2 Descriptives: Pre-reform and pre-part-time pension spell, constant variables, treatment and control groups

	Control	Treatment
Females, %	57.1	56.1
Living in the capital region***, %	27.3	23.5
Education, %		
Upper secondary educ.***	39.8	45.7
Short-cycle tertiary educ.*	26.7	24.5
Bachelor or equivalent	13.3	12.7
Master or equivalent**	17.5	15.2
Doctoral or equivalent**	2.7	1.8
Industry, %		
Manufacturing**	6.9	8.6
Retail	7.8	6.9
Professional service	6.2	6.1
Public administration*	10.5	9.2
Education***	15.2	11.3
Care taking	15.2	13.8
Occupations, %		
Managers	5.6	5.2
Professionals**	22.1	19.3
Technicians and associate professionals	20.8	19.3
Clerical support workers	12.9	14.2
Service and sales workers	11.5	11.7
Skilled agricultural, forestry and fishery workers	0.9	1.2
Craft and related trades workers***	7.7	10.3
Plant and machine operators and assemblers**	6.6	8.2
Elementary occupations*	11.2	9.8
Individuals	2 880	9 508

Notes: Only the biggest industries and occupations are listed. Manufacturing here combines pharmaceutical, electrical, machinery and vehicle industries. T-test for the difference in shares, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

ers and others. The complier group is defined as individuals who take-up part-time pension within a year when becoming eligible (reaching the cohort-specific eligibility age) while non-compliers take part-time pension spell at later ages. Tables 5.3 and 5.4 shows these descriptives. There are $\sim$ 35% of compliers in the sample. The compliers have more sickness days and drug purchases for both at the extensive and at the intensive margin. These differences are statistically significant at the 5% risk level apart from cerebrovascular diseases and diabetes. Women comply more with the statutory age limits and compliers tend to be represented more in the public sector.

Table 5.3 Descriptives on non-compliers and compliers, years 1995-1997

	Non-compliers	Compliers
Income and employment		
Wage income***, €	25 604.2 (13 340.9)	25 048.8 (11 631.8)
Net income, €	18 800.8 (18 890.6)	18 565.0 (10 448.1)
Pension income, €	243.5 (1 467.2)	230.6 (1 419.8)
Months in empl.**	11.62 (1.92)	11.57 (2.10)
Months in unempl.***	4.88 (3.68)	4.40 (3.48)
Health indicators		
Sickness absence days***	3.47 (15.78)	4.40 (17.77)
Any drug purchase***	0.55 (0.50)	0.57 (0.50)
Nr of purchases, any drugs***	2.69 (4.38)	2.91 (4.66)
Drug purchase for respiratory diseases*	0.24 (0.43)	0.25 (0.43)
Nr of purchases, resp. dis.	0.58 (1.71)	0.61 (1.75)
Drug purchase for circulatory diseases	0.20 (0.39)	0.20 (0.40)
Nr of purchases, circ. diseases	0.93 (2.38)	0.96 (2.42)
Drug for heart condition**	0.19 (0.39)	0.20 (0.40)
Nr of purchases, heart conditions	0.92 (2.38)	0.95 (2.41)
Cerebrovascular disease	0.01 (0.095)	0.01 (0.10)
Nr of purchases, cerebrovas. diseases	0.02 (0.27)	0.02 (0.27)
Musculo-skeletal disorder***	0.25 (0.43)	0.27 (0.44)
Nr of purchases, musculo-skeletal dis.***	0.49 (1.13)	0.54 (1.21)
Drugs for diabetes	0.02 (0.13)	0.02 (0.13)
Nr of purchases, diabetes	0.09 (0.78)	0.09 (0.82)
Mental illness drugs***	0.10 (0.30)	0.11 (0.32)
Nr of purchases, mental disease***	0.32 (1.60)	0.42 (1.85)
Individuals	34 336	17 961

Notes: Means with standard deviations in parentheses. All variables are measured within a year. Unit of measure for *drug purchases* is share of individuals with any purchases within a group and the *number of purchases* is measured in packages. *Sickness absence days* represent the absence days over 10 days. T-test for the differences in means, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.4 Empirical methodology

The main motivation for the part-time pension program was that it could reduce the risk of early labour market exit. Above all, it was expected that with the help of part-time work, disability-related exits would decrease, thus part-time pension scheme was based on the notion that work effort is related to health outcomes. This indicates that part-time retirement could also affect sickness absences and drug purchases beside the decrease in the number of disability spells.

The goal of this paper is two-fold: firstly, to study what kind of effects the change in the eligibility ages *(reform effects)* had on the health outcomes in the population of part-time pensioners, and secondly, what effects the take-up *(take-up effects)* of part-time pension had on the health outcomes of individu-

Table 5.4 Descriptives on non-compliers and compliers, years 1995-1997

	Non-compliers	Compliers
Females***, %	56	58
Living in the capital region***, %	24.6	22.3
Education, %		
Upper secondary educ.*	43.9	44.9
Short-cycle tertiary educ.	25.1	24.7
Bachelor or equivalent***	13.4	12.2
Master or equivalent**	15.5	16.4
Doctoral or equivalent**	2.17	1.81
Industry, %		
Manufacturing***	11.9	9,6
Retail*	4.45	3.98
Professional service*	6.24	5.79
Public administration	9.25	9.77
Education***	10.3	15.5
Care taking***	13.8	15.0
Occupations, %		
Managers***	5.88	4.14
Professionals***	19.1	21.8
Technicians and associate professionals***	20.2	18.3
Clerical support workers**	13.7	14.6
Service and sales workers	11.7	11.6
Skilled agricultural, forestry and fishery workers	1.17	1.10
Craft and related trades workers	9.56	10.0
Plant and machine operators and assemblers***	8.32	7.28
Elementary occupations**	9.62	10.5
Individuals	34 336	17 961

Notes: Only the biggest industries and occupations are listed. Manufacturing here combines pharmaceutical, electrical, machinery and vehicle industries. T-test for the differences in proportions, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

als. For the first question treatment and control groups are formed and their drug purchases and sickness absences are compared before and after the age limit reform in a difference-in-differences setting. These effects are intention-to-treat effects as not all individuals took the part-time pension at the age they were first entitled to it. The second question also exploits the eligibility ages but as an instrument for the actual retirement. This estimation strategy identifies the local average treatment effect (LATE) for the compliers who retire within the first year of becoming eligible for the part-time pension.

These research designs can answer the question of the relationship between health and labour market behaviour and work intensity from two perspectives. Firstly, the reform effects take the perspective of the policy maker and answer what kind of average effect the reform had and whether the reform succeeded at reducing the disability spells and improving the work ability.

The take-up effects have the perspective of the individual. As the IV design can reveal effects only on the subgroup, these can be different from those average effect of part-time pension where we could move all elderly workers to part-time pension.

5.4.1 Difference-in-differences setting

In July 1998 the eligibility age for part-time pension was lowered to 56. We can form a treatment group to be those eligible for part-time pension at the age of 56 in year 1998, i.e. individuals born between July 1942 and December 1946, and a control group, whose eligibility age remained at 58 and are so born between January to June 1940[7]. I use a regression difference-in-differences method to estimate the reform effects. I will estimate the following equation:

$$Y_{i,t} = \alpha + \beta_1 P_t + \beta_2 T_i + \beta_3 (P_t * T_i) + \mu_b + \gamma X_{it} + \epsilon_{it}, \tag{5.1}$$

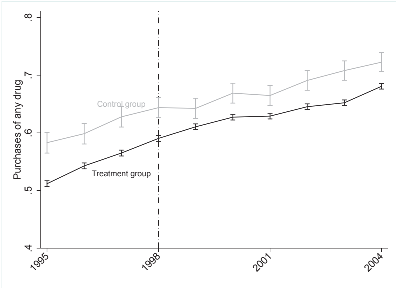
where $Y_{i,t}$ is the outcome for individual i in year t, P_t takes value one for years after 1998 and zero otherwise, T_i takes value one if the individual is in the treatment group and zero otherwise and $(P_t * T_i)$ is their interaction term. I also include a vector of control variables X_{it}, which are listed in the descriptive tables. Also cohort fixed effect μ_b are included. The coefficient β_3 is the estimated average effect of the reform.

The reform is identified in the difference-in-differences approach if certain assumptions are fulfilled. The main assumption is that there are parallel trends before the reform. Figure 5.2 shows the raw means and confidence intervals for the main outcomes. From these figures it is visible that the parallel trends hold clearly only for any drug purchases in the intensive margin. However, in the main regression specification I control for the background variables and show that the trend in the difference between the two groups is not statistically different from zero.

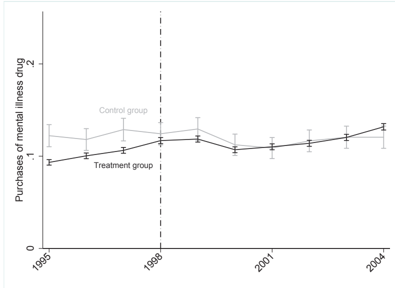
Another important assumption in the difference-in-differences setting is

[7]I argue that the reform was exogenous because none of the background material I have gone through preparing for this change in the legislation suggest anything of the kind that these specific cohorts are in need of special treatment due to their health outcomes. Also to my knowledge there are no common shocks that affect only the treatment group.

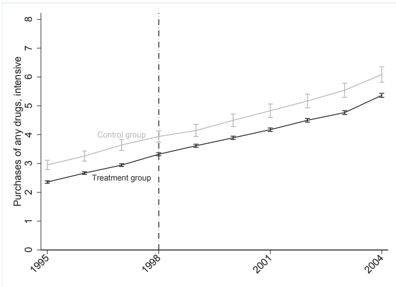
Figure 5.2 Main outcome variables, by year and treatment status.



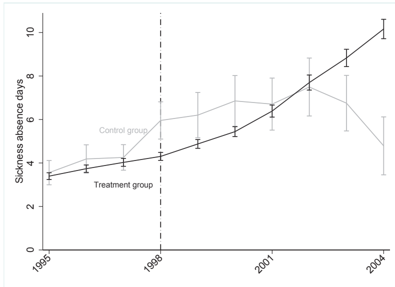
(a) Any drug purchases, extensive margin



(b) Mental illness drug purchases, extensive margin



(c) Any drug purchases, intensive margin



(d) Sickness absences days, over 10 days

Notes: Mean and 95% confidence intervals.

that the reform cannot affect the control group. In this setting the reform is based on birth cohorts so it is impossible to move from the control group to the treatment group. However, the reform can affect the behaviour of the control group. From the figure 5.2 we notice that there is a peculiar trend for the control group with respect to sickness absences days. During the reform year the sickness absence days increase in the control group. This can indicate that either something happened between 1998-2001 that affected only the control group or the increase in the sickness absence days for the control group is due to the reform. For example, if the control group considered the decrease in the age limit unjustifiable, this might have motivated some of them to seek possibilities to extend their sickness absence leaves. The data I have cannot solve this issue so the results with respect to the sickness absences should be interpreted with caution.

5.4.2 Fixed effects instrumental variables estimation

To determine the causal effects of working shorter hours on health is difficult because the work hours decision is endogenous to health which is unobserved. Individuals with worse health outcomes are more likely to reduce their work load or retire fully earlier. Estimating an OLS model of the form

$$Y_{i,t+k} = \beta PR_{i,t} + f(a_{i,t}) + X_{it}\gamma + \nu_{i,t}, \tag{5.2}$$

where health outcome $Y_{i,t+k}$ for individual i in a future period $t+k$ depends on part-time retirement status $PR_{i,t}$, age function $f(a_{i,t})$, the vector of the individual's characteristics X_{it} and an idiosyncratic error term $\nu_{i,t}$, is biased if part-time retirement decision and health are correlated. Another source of bias is the individual specific unobservable factors which can be correlated with both the part-time work decision and health. I tackle these problems with fixed-effects instrumental variable analysis.

The individual fixed-effects estimates compare the full-time job and part-time job status at the individual level and relate the change in the status to the changes in the health outcomes at the individual level. This strategy eliminates any unobserved time-invariant factors. To account for the unobserved heterogeneity and potentially endogenous decisions in the part-time retirement timing, I estimate within-two-stage least-squares and use the eligibility ages and the exogenous changes in these ages as an instrument for the part-time retirement decision[8]. That is, the estimating equation includes within transformation both in the first and the second stage[9].

The first stage regression takes the form of a linear probability model:

$$PR_{i,t} = \gamma 1[a_i = e_i] + f(a_{i,t}) + \rho_i + \epsilon_{i,t}, \tag{5.3}$$

where the part-time retirement, $PR_{i,t}$, takes values 0/1 indicating whether individual has taken part-time pension or not, e_i denotes the eligibility age

[8] The same kind of instrument is used for example by Bonsang et al. (2012), Kantarci and Kolodziej (2017), Lucifora and Vigani (2018) among others.

[9] In health and employment literature FE-IV estimation procedure has been used by Bonsang et al. (2012), Frijters et al. (2014), (Ahn, 2016), Kantarci and Kolodziej (2017), Cygan-Rehm and Wunder (2018) and Lucifora and Vigani (2018)

for individual i and $1[a_i = e_i]$, the instrument, is an indicator function taking value 1 if individual has reached the age of eligibility in the same year[10]. The γ measures the discontinuity in the probability to retire at this age. The unobserved individual specific time-invariant variables ρ_i are abolished with within-transformation.

For the instrumental variable to work, it is required that the instrument is relevant for the actual part-time pension decision and satisfies the exclusion restriction. Figure 5.3 shows the discontinuous change in the probability to take up part-time pension with respect to years from the eligibility. About a third of part-time pensioners take up the part-time pension during the year they become eligible. Figure 5.A2 in appendix shows the retirement timing separately for groups where the lowest eligibility age differs due to the reform taking place in 1998. While the reform made the programme more popular and the largest spike is at the age 56, the older cohorts also retire around the year they become eligible. The identification is based on this discontinuity in the proportion of individuals reducing work load through part-time retirement right at the eligibility age conditional on a polynomial function of age.

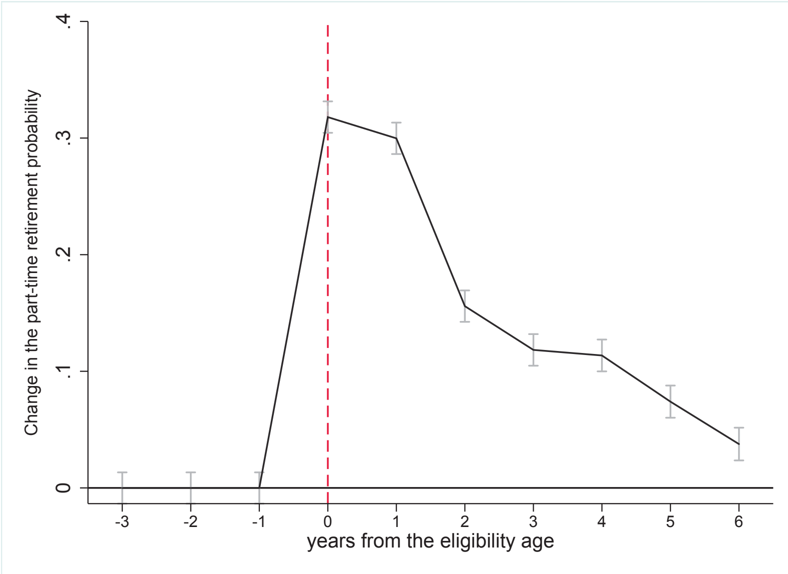
While the specific eligibility ages have a direct effect on the decision to take part-time pension, it is less probable that they have a particular effect on the outcome variables except through the part-time retirement[11]. The reverse causality, that the health would affect the instrument, would happen for example if the eligibility age is set to a certain age where health problems accumulate. I have gone through the government's proposition for the bill and the subsequent parliament discussion in order to see if these specific eligibility ages were chosen because of population's health conditions. This seems not to be the case.

The final estimation is based on the following second-stage fixed effects

[10]It is important to remember here that all individuals in the current sample take up part-time pension at some age. However, the timing differs and this instrumental variable research design is able to reveal the effects on the compliers who retire during the year becoming eligible. These are the local average treatment effects.

[11]This is supported by the fact that I have looked at the trends in the health-related factors for the total population and the eligibility ages for part-time pension are not anomalous.

Figure 5.3 Take-up of first part-time pension spell since becoming eligible



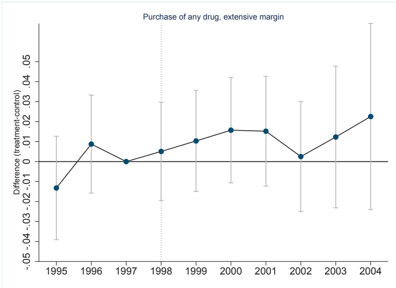
Notes: the estimates are based on a fixed effects model where years from the eligibility act as explanatory variables. The estimation is done for years 1995-2004 including the total part-time pensioners sample. The vertical lines represents 95% confidence intervals.

regression model

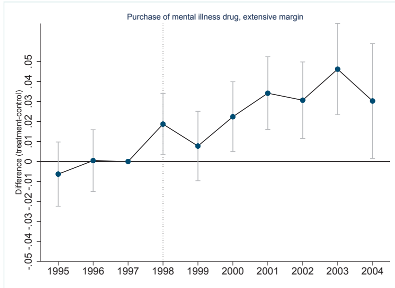
$$Y_{i,t+k} = \beta PR_{i,t} + f(a_{i,t}) + \mu_i + \nu_{i,t}, \tag{5.4}$$

where μ_i denotes the unobserved time-invariant heterogeneity and captures all time-invariant characteristics that are associated with both the decision to reduce work hours and outcome variables. The coefficient β is the parameter of primary interest and represents the impact that reducing work hours through part-time pension has on $Y_{i,t+k}$. The identification of β is driven by changes in the outcome variables for individuals whose part-time retirement decision is affected by the eligibility age.

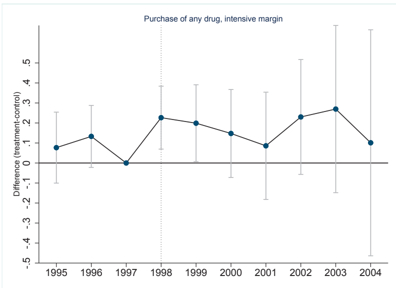
Figure 5.4 Difference in outcomes between the treatment and the control group, with controls.



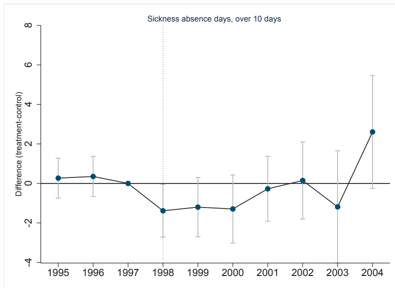
(a) Any drug purchases, extensive margin



(b) Mental illness drug purchases, extensive margin



(c) Any drug purchases, intensive margin



(d) Sickness absences days, over 10 days

Notes: Coefficient for treatment group and 95% confidence intervals. Sickness absence days conditional on working. Standard errors are clustered at individual level.

5.5 Results

5.5.1 Reform effects

I begin by estimating an event-study model in order to graph the differences over time between the treatment and the control groups. The estimated equation is

$$Y_{i,t} = \alpha_i + \sum_{t=1995}^{2004} \beta_t T_i + \gamma X_{it} + \mu_b + \epsilon_{it}, \tag{5.5}$$

where β_t is the estimated difference between the control and treatment group in each year conditional on the vector of control variables and cohort fixed effects.

The figure 5.4 shows the results from this estimation with respect to the main outcomes. The figures without controls are presented in the appendix (figure 5.A3). The coefficients pre-reform are close to zero and statistically we cannot reject the null hypothesis that the two groups do not differ. We observe from these figures that the drug purchases increase in the treatment group after the age limit was lowered, most clearly for mental illness drugs. We also observe that there is a reduction in the difference in the sickness absence days. However, these estimates are rather imprecise.

In the figures the confidence intervals are quite large so I next present simple difference-in-differences estimates where years are pooled together. I study the effects until the year 2001 and so I have 3 years before the reform and 3 years after. Beside the sickness absence days, all coefficients are positive indicating that the reform increased the purchases of drugs by the treatment group. One mechanism could be that individuals have more leisure during the part-time pension and this time can be spent for example by going to see a medical professional. However, as I have only certain type of register data I cannot test different mechanisms. Also it is noticeable that the coefficients are imprecise when we include control variables, beside the results for the mental illness drug purchases. For mental illness drug purchases the coefficient is large and statistically significant also after including controls. After adding controls, the coefficient on the sickness absences also becomes significant and indicates that the reform had on average a reduction of 1.25 days in the sickness absence spells for the treatment group. However, this is driven by the change in the sickness absence of the control group and the meaning of this result is unclear.

One of the goals set for the part-time pension scheme and the lowering of the eligibility age was to increase old-age retirement age. Unfortunately this is impossible to study because of the pension reform in 2005. The pension reform in the year 2005 lowered the retirement age for the cohorts born after 1942 and also affected the financial incentives. This reform intervenes with the current study design and so I have focused on the period before the pension reform. However, I have analysed the effect on the probability of retiring through some early-retirement pension scheme (typically means disability pension or individual early retirement) before the year 2005. As shown in table 5.6, the early exit actually increased in the treatment group by 5,5 per-

Table 5.5 Difference-in-differences (DD) results on main outcomes

	Any drug purchase	Mental ill. drug purchase	Any drug purchase	Mental ill. drug purchase
DD estimate	0.0269***	0.0181***	0.0163	0.0231***
	(0.0067)	(0.0047)	(0.0094)	(0.0065)
Cohort & Gender	YES	YES	YES	YES
Controls	NO	NO	YES	YES
R^2	0.027	0.0141	0.0852	0.0456
Observations	236 406	236 406	132 863	132 863
Individuals	39 422	39 422	22 449	22 449

	Any drug purchase, int.	Sickness abs.	Any drug purchase, int.	Sickness abs.
DD estimate	0.0296	-0.7394	0.0937	-1.2564**
	(0.0711)	(0.3803)	(0.0905)	(0.4753)
Cohort & Gender	YES	YES	YES	YES
Controls	NO	NO	YES	YES
R^2	0.021	0.033	0.073	0.102
Observations	236 406	236 065	132 863	132 792
Individuals	39 422	39 422	22 449	22 449

Notes: Years in the estimation are 1995-2001. Cluster robust standard errors are in parentheses (clustered at individual level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

Table 5.6 Probability to retire through early-pension scheme

	Early retirement
Treatment group	0.0555***
	(0.0087)
Cohort& Gender	YES
Controls	YES
Observations	213 479
Individuals	22 482

Notes: Cluster robust standard errors are in parentheses (clustered at individual level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

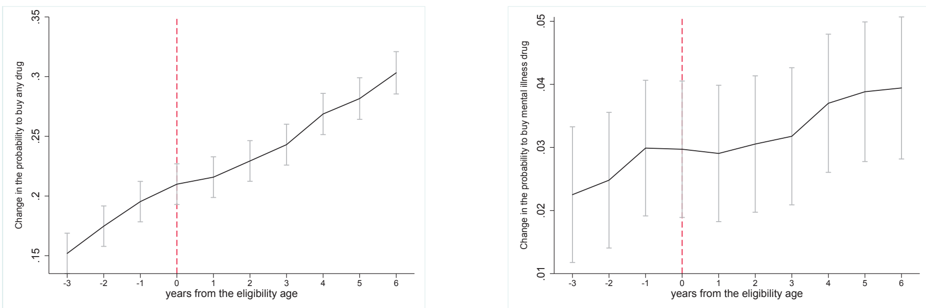
centage points compared to the control group. The longer period of reduced work load did not lead to later labour market exits[12].

[12] In the appendix figure 5.A4 shows further evidence that the part-time retirement scheme was not successful in increasing the age of labour market exits. This descriptive evidence shows average retirement ages by birth months for cohorts 1946 and 1947. The cohorts differ with regards to their part-time retirement eligibility age while their old-age statutory age was the same. Cohort 1946 had the part-time pension age limit at 56 and for cohort 1947 the part-time retirement age was 58. So there is a discontinuity in the part-time pension age and observing the average retirement age for both sides of the threshold reveal that for cohort 1946 the average retirement age is clearly lower. I have also checked other observable background variables for whether they differ around the threshold. This seems not to be the case but I cannot do a formal regression discontinuity design analysis as above the threshold there are only approximately 2000 observations over the whole 1947 birth cohort.

5.5.2 Average effect on the compliers

In this section I study the short-term effects of taking-up part-time pension. Here the outcomes are measured one year after taking up part-time retirement. Figure 5.5 shows the change in the probability of purchasing a positive amount of any medicine or mental illness drug with respect to years since being eligible for the part-time pension. The figure points out that there is a small change in the slope after being eligible for part-time pension. This change is stronger for the mental illness drugs and is observed already in the year of retiring. These changes are not long-lasting. As a placebo test I graphed the same figures for the non-part-time pensioners sample (appendix figure 5.A5). For any drug purchases there is no clear change in the slope so it seems that the decrease in the probability to buy drugs is unique to the part-time pensioners. However for mental illness drugs the picture is blurrier and there seems to be more age-driven changes in the purchases.[13]

Figure 5.5 Change in the probability to buy any medicine (left figure) or mental illness drug (right figure) during a year by distance from the eligibility age.



Notes: the estimates are based on a fixed effects model where years from the eligibility act as explanatory variables excluding non-part-time pensioners' sample. The vertical lines represents 95% confidence intervals. Years in the estimation are 1995-2004.

While the previous observations tell the average outcomes within the total population of part-time pensioners based on the eligibility age, running the

[13]Table 5.A6 in the appendix shows the fixed effect model estimates based on take-up. The change in drug purchases is clear in the year of taking the part-time pension. These are effectively the fixed effects results without taking into account the endogeneity which is also presented in the regression tables for comparison.

regression model based on the equations 5.3 and 5.4 gives the estimates of the effect within the group of compliers. Table 5.7 presents the first set of these results estimated by the two-stage least squares within estimator.

The first column displays the coefficients from the first-stage regression (eq. 5.3). The instrument, that is the eligibility age, has a large and highly significant effect on the probability of having reduced working hours (taken part-time pension). The first stage F-statistic is 43.31 which is well above the value of 10 which is commonly used as a cutoff value for a good instrument (Staiger and Stock, 1997).

The effect of transition to part-time retirement on the probability to purchase any drug is negative and statistically significant (column 2). The estimate indicates that the working hour reduction due to part-time pension lead to 2.8 percentage point lower drug purchases on average within the compliers. This estimate is somewhat stronger than in the model where exogeneity of retirement is assumed (column 3). The relative effect is a 4.9 % reduction (compared to the sample average before part-time retirement). For the probability to buy mental illness drugs the effects are much smaller. Table 5.A4 in the appendix shows the estimation results by drug category. By subcategory the estimated coefficients from the IV-specification are strongest for the respiratory and musculo-skeletal diseases drugs at the 5% or 10% risk level, respectively.

Tables 5.8 and 5.9 show the results for men and women separately. There are noticeable gender differences. In the short-term, the part-time retirement leads women to purchase 3.6 percentage point less (any) drugs while for men the effect is 1.7 percentage points and this difference is statistically significant. These estimates mean that there is a strong 5.8% relative effect for women while the relative effect is slightly smaller for men being 3.5%. Also it is noticeable that for women there is also a statistically significant and relevant reduction of 1.3 percentage points for the purchases of mental illness drugs.

For robustness, I have also explored the effects of the choice of the functional form for the age term. I tested linear, quadratic, cubic and quartic functional forms. The part-time pension coefficients are quite insensitive to the functional form, however the standard errors increase quite a lot in the cubic and quartic specification. Also for the cubic and quartic functional form the

Table 5.7 Number of purchases of any drug or mental illness drug and part-time retirement

	Taken part-time pension at t First stage (1)	Any drug IV-FE (2)	Any drug linear-FE (3)	Mental illness drug IV-FE (4)	Mental illness drug linear-FE (5)
eligible	0.2174***				
	(0.0555)				
age	0.2088*	0.0305***	0.0264*	0.0334***	0.0316***
	(0.0839)	(0.0078)	(0.0085)	(0.0040)	(0.0045)
age^2	-0.0017*	-0.0001	-0.0001	-0.0003***	-0.0003***
	(0.0008)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
part-time retirement, PR		-0.0280***	-0.0114***	-0.0090***	-0.0019
		(0.0067)	(0.0017)	(0.0026)	(0.0010)
Constant			-0.5552		-0.8414***
			(0.2390)		(0.1253)
Within R2			0.019		0.002
F-statistic		43.31		43.31	
Observations	521 155	521 155	521 156	521 155	521 156

Notes: Years in the estimation are 1995-2004. Regressions include year dummies. Health outcomes measured in period t+1. Cluster robust standard errors are in parentheses (clustered on birth cohort level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

Table 5.8 Purchase of any amount of drug and part-time retirement, men

	Taken part-time pension at t First stage (1)	Any drug IV-FE (2)	Any drug linear-FE (3)	Mental illness drug IV-FE (4)	Mental illness drug linear-FE (5)
eligible	0.2052***				
	(0.0294)				
age	0.1454**	0.0210***	0.0196***	0.0102	0.0098
	(0.0503)	(0.0035)	(0.0033)	(0.0057)	(0.0074)
age^2	-0.0013**	-0.0000	0.0000	-0.0001	-0.0001
	(0.0005)	(0.0000)	(0.0000)	(0.0001)	(0.0001)
part-time retirement, PR		-0.0171*	-0.0109**	-0.0024	-0.0007
		(0.0076)	(0.0024)	(0.0061)	(0.0010)
Constant			-0.5357***		-0.3007
			(0.0919)		(0.2034)
Within R^2			0.027		0.004
F-stat		48.66		48.66	
observations	226 878	226 878	226 878	226 878	226 878

Notes: Years in the estimation are 1995-2004. Regressions include year dummies. Health outcomes measured in period t+1. Cluster robust standard errors are in parentheses (clustered on birth cohort level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

Table 5.9 Purchase of any amount of drug and part-time retirement, women

	Taken part-time pension at t First stage (1)	Any drug IV-FE (2)	Any drug linear-FE (3)	Mental illness drug IV-FE (4)	Mental illness drug linear-FE (5)
eligible	0.2186***				
	(0.0347)				
age	0.1637**	0.0376**	0.0314	0.0509***	0.0480***
	(0.0554)	(0.0136)	(0.0153)	(0.0046)	(0.0049)
age^2	-0.0014**	-0.0002	-0.0002	-0.0004***	-0.0004***
	(0.0005)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
part-time retirement, PR		-0.0361***	-0.0122***	-0.0138*	-0.0029
		(0.0081)	(0.0021)	(0.0061)	(0.0018)
Constant			-0.5640		-1.2477***
			(0.4298)		(0.1325)
Within R^2			0.014		0.002
F-stat		39.79		39.79	
observations	294 277	294 277	294 278	294 277	294 278

Notes: Years in the estimation are 1995-2004. Regressions include year dummies. Health outcomes measured in period t+1. Cluster robust standard errors are in parentheses (clustered on birth cohort level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

coefficients for ages are insignificant. I conclude that the quadratic specification I have used is satisfactory to capture the non-linearities in age.

In the intensive margin, the effects of part-time retirement on the amount (in packages) of purchased drugs are shown in table 5.10 for the pooled sample and also for both of the genders separately. On average there is a quarter of package reduction for the compliers due to the part-time retirement. The gender decomposition shows that this effect comes from the women's purchases while for men working hours reduction does not have any effect. The women's point estimate -0.39 translates to 12% reduction of medicine use. Most of this reduction comes from the reduction in the purchases of drugs for musculo-skeletal and circulatory diseases[14].

The distribution of sickness absence days is highly right-skewed. Within a year approximately 85 percent of the individuals have zero days of long absenteeism[15] while there are very few with extremely long absenteeism. Over the observation years, however, there are only about a quarter of the sample who do not have any sickness benefit spells. As I am not familiar with a non-linear estimator that would take into account the unobservable heterogeneity and endogenous covariate, I transform the sickness absence days data

[14]Results available upon request from the author.

[15]As mentioned in the section 5.3.2 only absences exceeding 10 days are covered by the social insurance and are included in the data.

Table 5.10 Amount of drug packages purchased and part-time retirement

	Overall					Men					Women				
	PR	any drug		mental illness drug		PR	any drug		mental illness drug		PR	any drug		mental illness drug	
	First stage	IV-FE	linear-FE	IV-FE	linear-FE	First stage	IV-FE	linear-FE	IV-FE	linear-FE	First stage	IV-FE	linear-FE	IV-FE	linear-FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)		
eligible	0.2127***					0.2052***					0.2186***				
	(0.0323)					(0.0294)					(0.0347)				
age	0.1559**	0.0097	-0.0365	0.1360***	0.1347***	0.1454**	-0.2354*	-0.2431	0.0395	0.0492	0.1637**	0.1945**	0.1191	0.2093***	0.1994***
	(0.0532)	(0.0772)	(0.0777)	(0.0197)	(0.0199)	(0.0503)	(0.1110)	(0.1111)	(0.0210)	(0.0227)	(0.0554)	(0.0668)	(0.0678)	(0.0263)	(0.0270)
age^2	-0.0014**	0.0030***	0.0034**	-0.0011***	-0.0011***	-0.0013**	0.0054***	0.0055***	-0.0002	-0.0003	-0.0014**	0.0012	0.0018*	-0.0017***	-0.0016***
	(0.0005)	(0.0007)	(0.0007)	(0.0002)	(0.0002)	(0.0005)	(0.0010)	(0.0010)	(0.0002)	(0.0002)	(0.0005)	(0.0006)	(0.0006)	(0.0002)	(0.0002)
Part-time ret., PR		-0.2669***	-0.0824**	-0.0193	-0.0142		-0.0893	-0.0571**	0.0352	-0.0057		-0.3956***	-0.1052*	-0.0590**	-0.0209*
		(0.0484)	(0.0234)	(0.0196)	(0.0070)		(0.0690)	(0.0143)	(0.0192)	(0.0063)		(0.0647)	(0.0326)	(0.0216)	(0.0088)
Constant			-4.1385		-3.7578***			0.5606		-1.5339			-7.6628**		-5.4368***
			(2.1278)		(0.5576)			(3.0865)		(0.6560)			(1.8242)		(0.7532)
Within R2			0.090		0.002			0.102		0.003			0.082		0.002
F-stat		43.31		43.31			48.66		48.66			39.79		39.79	
Observations	521 155	521 155	521 156	521 155	521 156	226 878	226 878	226 878	226 878	226 878	294 277	294 277	294 278	294 277	294 278

[a] Notes: Years in the estimation are 1995-2004. Regressions include year dummies. Health outcome is the number of drug package purchase within a year and it is measured in period t+1. Cluster robust standard errors are in parentheses (clustered on birth cohort level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

Table 5.11 Probability to have over 10 days of sickness absences and part-time retirement

	Overall			Men			Women		
	First stage	IV-FE	linear FE	First stage	IV-FE	linear FE	First stage	IV-FE	linear FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
eligible	0.2122***			0.2049***			0.2179***		
	(0.0318)			(0.0288)			(0.0341)		
age	0.1314*	0.1561***	0.1443***	0.1187*	0.1185***	0.1094**	0.1408*	0.1842***	-0.1705***
	(0.0547)	(0.0127)	(0.0131)	(0.0516)	(0.0111)	(0.0115)	(0.0570)	(0.0138)	(0.0141)
age^2	-0.0011*	-0.0014***	-0.0013**	-0.0010*	-0.0011***	-0.0010***	-0.0012*	-0.0017***	-0.0016***
	(0.0005)	(0.0001)	(0.0001)	(0.0005)	(0.0001)	(0.0001)	(0.0005)	(0.0001)	(0.0001)
Part-time retirement, PR		-0.0691***	-0.0177***		-0.0509***	-0.0084*		-0.0822***	-0.0253***
		(0.0084)	(0.0031)		(0.0101)	(0.0024)		(0.0103)	(0.0042)
Constant			-3.7654***			-2.8948***			-4.4187***
			(0.3623)			(0.5211)			(0.7375)
Within R2			0.007			0.006			0.012
F-stat		44.61			50.43			40.81	
Observations	511 973	511 973	511 973	222 871	222 871	222 871	289 102	289 102	289 102

Notes: Years in the estimation are 1995-2004. Regressions control for year and previous sickness absence days while the outcome (probability of having sickness benefit spells) is measured in t+1. Cluster robust standard errors are in parentheses (clustered on birth cohort level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

to a binary variable indicating whether individual experience any duration of sickness absence. These results are presented in the table 5.11.

The effect on probability of long sickness absence is negative and highly significant. The part-time retirement leads to a 6.9 percentage point reduction in this probability. The effect is much larger than in the linear fixed effects specification. The reason lying behind this direction of differences is probably that the IV-FE identify the effect on the compliers while in the model where the retirement decision is considered to be exogenous the estimated effect is average effect in the sample. There is a statistically significant difference between men and women. Women's probability reduces by 8.2 percentage points while for men the effect is 5 percentage points.

As with the sickness absence the labour market exits are also modelled as a binary variable. Here I am interested in how reducing the work hours affects one's probability to exit permanently out from the labour market via an early retirement scheme. The outcome variable takes the value 1 if the individual has an early exit. I also add sickness absence days (within a year) as an explanatory variable since retirement literature has shown that health is one primary explanatory variable in the retirement decision. Table 5.12 shows the estimation results.

The first 3 columns show the results based on the pooled sample while the next 6 columns present results separately for gender. The point estimate -0.051 suggests that there is a significant reduction in the probability to tran-

Table 5.12 Probability of early labour market exit and part-time retirement

	Overall			Men			Women		
	First stage (1)	IV-FE (2)	linear FE (3)	First stage (4)	IV-FE (5)	linear FE (6)	First stage (7)	IV-FE (8)	linear FE (9)
eligible	0.2122***			0.2049***			0.2179***		
	(0.0318)			(0.0289)			(0.0342)		
age	0.1315*	-0.0915***	-0.1009***	0.1187*	-0.0974***	-0.0618**	0.1410*	-0.0872***	-0.0562*
	(0.0547)	(0.0138)	(0.0127)	(0.0517)	(0.0115)	(0.0163)	(0.0571)	(0.0164)	(0.0194)
age^2	-0.0011*	0.0009***	0.0010***	-0.0010*	0.0009***	0.0006**	-0.0012*	0.0008***	0.0005*
	(0.0005)	(0.0001)	(0.0001)	(0.0005)	(0.0001)	(0.0002)	(0.0005)	(0.0002)	(0.0002)
Part-time retirement, PR		-0.0518***	-0.0108***		-0.0565***	-0.0082*		-0.0485***	-0.0101**
		(0.0142)	(0.0016)		(0.0161)	(0.0025)		(0.0134)	(0.0021)
Within R2			0.091			0.029			
F-stat		44.51			50.42			40.66	
Observations	512 126	512 126	512 127	222 934	222 934	222 934	289 192	289 192	289 192

[a] Notes: Years in the estimation are 1995-2004. Regressions control for year and previous sickness absence days while the outcome (probability of having sickness benefit spells) is measured in t+1. Cluster robust standard errors are in parentheses (clustered on birth cohort level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

sit out of the labour market via early exit. An earlier study by Kyyrä (2015) estimates the eligibility effects for the same observation period. His findings with respect to part-time pension is that the eligibility caused a reduction in the probability to transit to unemployment (especially for public sector workers) but no statistically significant results are found for disability pension. The difference with the current study is that he considers the eligibility effect on the overall population while here I account the take-up of part-time pension and identify the work reduction effect on the complying individuals. In this sense these studies complement one another.

While for the sickness absences and the drug purchases, women had larger reductions than men, for the early market exits the effect is stronger for men and this difference is also statistically significant. The probability of women to exit via early route is 4.8 percentage points lower while for men the figure is 5.6 percentage points.

5.6 Conclusion

This paper studied the effect of working hours reduction on health-related factors and early labour market exits for elderly workers. The working hours reduction is studied in the context of the part-time pension which provides a good setting as this pension scheme certainly affected the hours worked but had a modest impact on the disposable income and future pension rights and generally did not change the work community. I studied first the effect of the change in the eligibility ages and secondly the effects of take-up of part-time pension accounting for individual heterogeneity and the endogeneity of the working hours decision.

The estimation results with respect to the reform effect are positive but imprecise on the drug purchases. There is no evidence found for the hypothesis that the part-time pension would prolong the work careers. The take-up of part-time retirement decreased the drug purchases and the probability of being on long sickness absence and these effects were larger for women. There seems to be a direct work-related aspect here as purchases of respiratory and musculo-skeletal diseases decreased the most. Part-time pension also reduced the probability of an early exit from the labour market. However, these results are local average treatment effects and are not as such extendible to the larger population.

The descriptive statistics showed that the compliers are negatively selected with respect to their health. However, compared to the general public we observed that the complying part-time pensioners are better educated and have better health outcomes. In this respect these results can be thought of as lower bounds or at least we cannot state that a work hours reduction would not be beneficial for some other subpopulation. It would be worthwhile to match the part-time pension sample to the other employees to study the effects in a larger population.

The limitation of the study is that I cannot explore the different mechanisms behind observed patterns. Increasing leisure time can affect life habits and health in various way. For example, in Ahn (2016) it is shown that a shortened work-week increases the likelihood of regular exercise and decreases the likelihood of smoking. Regular exercise or other personal investments on one's own health could be behind the take-up effects. On the other hand, more leisure available could lead to more doctor visits and more prescribed medication which can be either preventive care or curing. This could be behind an observation that drug purchases increase.

Lastly, it should be mentioned that the part-time pension system studied in this paper was abolished in the pension reform in 2017. The major reason for this was that the past system was expensive and did not treat individuals in the same manner. A replacement scheme was created in which there are no work-related conditions assigned for claiming part of the earned pension rights beforehand. In the future, finding a good research design to study the working hours effects for the elderly population will be slightly more difficult.

Bibliography

Ahn, T. (2016). Reduction of Working Time: Does It Lead to a Healthy Lifestyle? *Health Economics*, 25(8):969–983.

Behncke, S. (2012). Does Retirement Trigger Ill Health? *Health Economics*, 21(3):282–300.

Bonsang, E., Adam, S., and Perelman, S. (2012). Does Retirement Affect Cognitive Functioning? *Journal of Health Economics*, 31(3):490–501.

Cygan-Rehm, K. and Wunder, C. (2018). Do Working Hours Affect Health? Evidence from Statutory Workweek Regulations in Germany. *Labour Economics*, 53:162–171.

Dave, D., Rashad, I., and Spasojevic, J. (2008). The Effects of Retirement on Physical and Mental Health Outcomes. *Southern Economic Journal*, 75(2):497–524.

Frijters, P., W. Johnston, D., and A. Shields, M. (2014). The Effect of Mental Health on Employment: Evidence from Australian Panel Data. *Health Economics*, 23:1058–1071.

Hagen, J. (2018). The Effects of Increasing the Normal Retirement Age on Health Care Utilization and Mortality. *Journal of Population Economics*, 31:193–234.

Hallberg, D., Johansson, P., and Josephson, M. (2015). Is an Early Retirement Offer Good for your Health? Quasi-Experimental Evidence from the Army. *Journal of Health Economics*, 44(C):274–285.

Hernaes, E., Markussen, S., Piggott, J., and Vestad, O. L. (2013). Does Retirement Age Impact Mortality? *Journal of Health Economics*, 32(3):586 – 598.

Imbens, G. and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–75.

Kantarci, T. and Kolodziej, I. (2017). Effects of Working Part-Time and Full-Time on Physical and Mental Health in Old Age in Europe. Mimeo.

Kivimäki, M., Jokela, M., and et. al. (2015). Long Working Hours and Risk of Coronary Heart Disease and Stroke: a Systematic Review and Meta-Analysis of Published and Unpublished Data for 603 838 Individuals. *Lancet*, 386:1739–46.

Kyyrä, T. (2015). Early Retirement Policy in the Presence of Competing Exit Pathways: Evidence from Pension Reforms in Finland. *Economica*, 82:46–78.

Leinonen, T., Laaksonen, M., Chandola, T., and Martikainen, P. (2016). Health as a Predictor of Early Retirement Before and After Introduction of a Flexible Statutory Pension Age in Finland. *Social Science & Medicine*, 158:149–157.

Lucifora, C. and Vigani, D. (2018). Health Care Utilization at Retirement: The Role of the opportunity Cost of Time. *Health Economics*, 27(12):2030–2050.

Rohwedder, S. and Willis, R. (2010). Mental Retirement. *Journal of Economic Perspectives*, 24(1):119–38.

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557.

Takala, M. (2004). Osa-aikaeläkeläisten eläkkeet ja ansiot. *Eläketurvakeskuksen monisteita 47*.

Taris, T. W., Ybema, J. F., Beckers, D. G. J., Verheijden, M. W., Geurts, S. A. E., and Kompier, M. A. J. (2011). Investigating the Associations among Overtime Work, Health Behaviors, and Health: A Longitudinal Study among Full-time Employees. *International Journal of Behavioral Medicine*, 18(4):352–360.

Virtanen, M., Heikkilä, K., Jokela, M., Ferrie, J., Batty, D., Vahtera, J., and Kivimäki, M. (2012). Long Working Hours and Coronary Heart Disease: A Systematic Review and Meta-Analysis. *American Journal of Epidemiology*, 176(7):586–96.

Virtanen, M., Jokela, M., and al. (2018). Long Working Hours and Depressive Symptoms: Systematic Review and Meta-Analysis of Published Studies and Unpublished Individual Participant Data. *Scandinavian Journal of Work, Environment & Health*, May 1;44(3):239–250.

Appendix

Table 5.A1 Classification of drug data

	ATC-codes
physical conditions	
Any disease	all ATCs
Cancer	L01
Respiratory diseases	R
Circulatory diseases	C
Heart related diseases	C01, C02, C03, C04, C07, C08, C09, C10
Cerebrovascular diseases	B01
Musculo-skeletal disorders	M01, M03, A03D, M02A, A03EA
Diabetes	A10
mental conditions	
Any mental illness	N05A, N05B, N05C, N06A
Anti-psychotics	N05A
Anxiolytics	N05B
Hypnotics and sedatives	N05C
Antidepressants	N06A

Sources: Hagen (2018), Leinonen et al. (2016), World Health Organization

Table 5.A2 Descriptives, years 1995-1997, part-time pensioners sample and comparison group

	Part-time pensioners	Others
Income and employment		
Labour income, €	25 412.9 (12 780.9)	20 095.2 (15 195.4)
Pension income, €	239.1 (1 451.1)	360.0 (1 652.5)
Net income, €	18 719.74 (16 485.12)	14 648.44 (15 583.30)
Months in empl.	11.60 (1.98)	8.61 (5.10)
Months in unempl.	4.75 (3.63)	7.99 (3.87)
Health indicators		
Sickness absence days	3.78 (16.50)	6.37 (30.38)
Any drug purchase	0.55 (0.50)	0.53 (0.50)
Nr of purchases, any drugs	2.77 (4.48)	3.09 (5.59)
Drug purchase for respiratory diseases	0.24 (0.43)	0.21 (0.41)
Nr of purchases, resp. dis.	0.59 (1.72)	0.58 (1.93)
Drug purchase for circulatory diseases	0.20 (0.40)	0.20 (0.40)
Nr of purchases, circ. diseases	0.94 (2.40)	0.99 (2.56)
Heart condition	0.19 (0.40)	0.20 (0.40)
Nr of purchases, heart conditions	0.93 (2.39)	0.98 (2.55)
Cerebrovascular disease	0.01 (0.10)	0.01(0.11)
Nr of purchases, cerebrovas. diseases	0.02 (0.27)	0.03 (0.31)
Musculo-skeletal disorder	0.26 (0.44)	0.24 (0.42)
Nr of purchases, musculo-skeletal dis.	0.51 (1.16)	0.51 (1.33)
Diabetes	0.02 (0.13)	0.02 (0.15)
Nr of purchases, diabetes	0.09 (0.80)	0.12 (0.93)
Mental illness drug	0.11 (0.30)	0.12 (0.33)
Nr of purchases, mental disease	0.36 (1.70)	0.57 (2.59)
Individuals	52 297	154 181

Means with standard deviations in parentheses. The first column shows descriptive statistics for years 1995-1997 for individuals who take-up part-time pension some point after year 1998. The second column shows the descriptives for comparison group who have not taken the part-time pension. All variables are measured within a year. Unit of measure for *drug purchases* is share of individuals with any purchases within a group and the *number of purchases* is measured in packages. *Sickness absence days* represent the absence days over 10 days.

Table 5.A3 Descriptives, years 1995-1997, part-time pensioners sample and comparison group, constant background variables

	Part-time pensioners	Others
Females, %	56	56
Living in the capital region, %	23.8	18.5
Education		
Upper secondary, %	44.2	55.0
Tertiary education, %	25.0	24.0
Bachelor, %	13.0	10.3
Master, %	15.8	9.2
Doctoral, %	2.0	1.5
Industry		
Manufacturing, %	8,3	4,8
Retail, %	4.3	4.4
Professional service, %	6.1	3.7
Public administration, %	9.4	4.6
Education, %	12.1	5.7
Care taking, %	14.2	11.9
Occupations		
Managers, %	5.3	4.8
Professionals, %	20.0	13.6
Technicians and associate prof., %	19.5	17.8
Clerical support workers, %	14.0	10.2
Service and sales workers, %	11.7	15.0
Skilled agricultural workers, %	1.1	8.4
Craft and related trades workers, %	9.7	9.4
Plant and machine operators and assemblers, %	8.0	10.1
Elementary occupations, %	9.9	9.0
Individuals	52 297	154 181

Notes: Only the biggest industries and occupations are listed. *Manufacturing* here combines pharmaceutical, electrical, machinery and vehicle industries. The first column shows the part-time pensioners sample and the second column shows the comparison group generated from the similar gender and birth cohort distribution as the part-time pensioners.

Table 5.A4 Purchase of any amount of drug and part-time retirement, by drug category

	First stage IV-FE	Respiratory dis. IV-FE	Respiratory dis. linear-FE	Heart dis. IV-FE	Heart dis linear-FE	Cerebro-vas. IV-FE	Cerebro-vas. linear-FE	Musculo-ske. IV-FE	Musculo-ske. linear-FE	Diabetes IV-FE	Diabetes linear-FE
eligible	0.2174***										
	(0.0555)										
age	0.2088*	0.0280***	0.0265**	0.0005	0.0012	-0.0117***	-0.0119**	0.0629***	0.0585***	-0.0045	-0.0051
	(0.0839)	(0.0051)	(0.0052)	(0.0060)	(0.0059)	(0.0031)	(0.0032)	(0.0043)	(0.0051)	(0.0023)	(0.0023)
age^2	-0.0017*	-0.0003***	-0.0003***	0.0003***	0.0003**	0.0001***	0.0001**	-0.0006***	-0.0005***	0.0001***	0.0001**
	(0.0008)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
part-time retirement, PR		-0.0162**	-0.0103***	-0.0003	-0.0032**	-0.0021	-0.0010	-0.0254*	-0.0079*	-0.0019**	0.0005
		(0.0059)	(0.0013)	(0.0041)	(0.0008)	(0.0014)	(0.0007)	(0.0101)	(0.0028)	(0.0006)	(0.0003)
Within R^2			0.012		0.117		0.009		0.001		0.021
F-stat		43.31087		43.31087		43.31087		43.31087		43.31087	
observations	521 155	521 155	521 156	521 155	521 156	521 155	521 156	521 155	521 156	521 155	521 156

[a] Notes: Years in the estimation are 1995-2004. Regressions include year dummies. Health outcomes measured in the next period. Cluster robust standard errors are in parentheses (clustered on birth cohort level). *, **, and *** indicate statistical significance at the 0.1, 0.05, and 0.01 levels, respectively.

Figure 5.A1 Length of the part-time pension spell

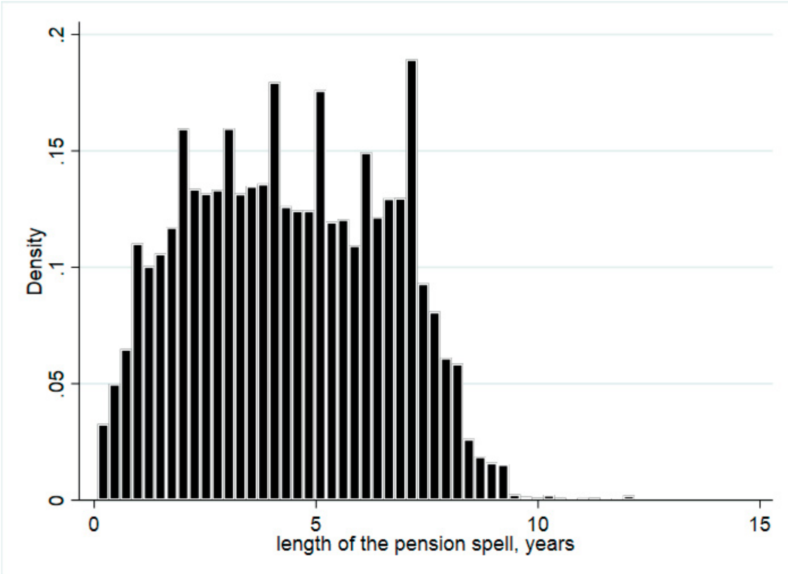


Figure 5.A2 Age distribution of part-time retirement by eligibility ages. The age at the starting day of the first part-time pension spell.

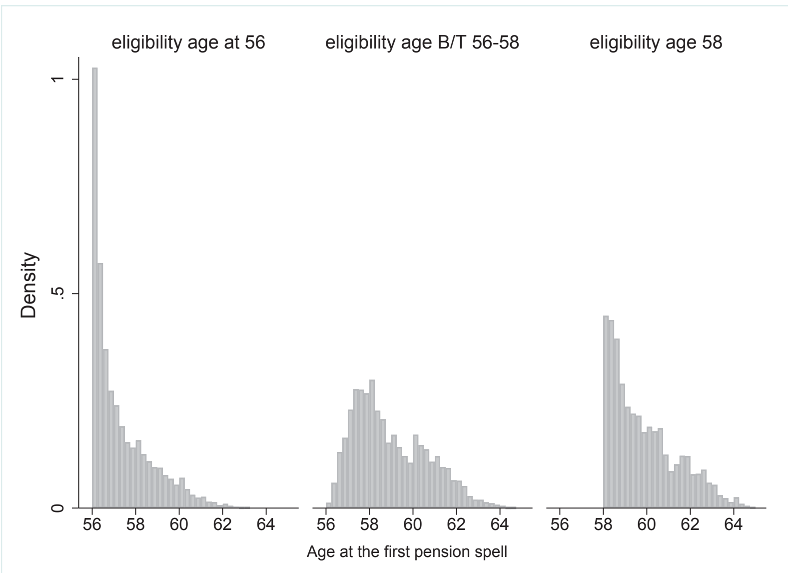
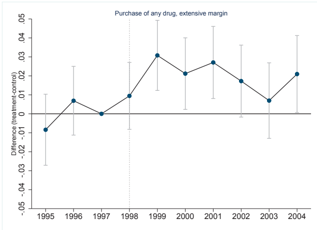
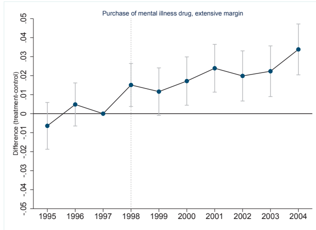


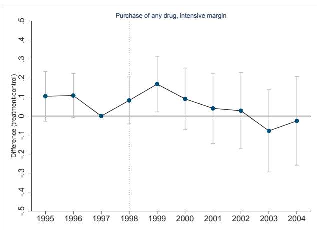
Figure 5.A3 Difference in outcomes between the treatment and the control group, without controls.



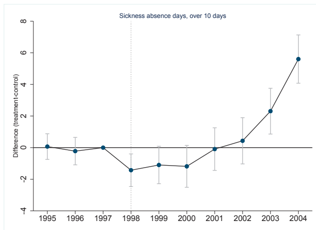
(a) Any drug purchases, extensive margin



(b) Mental illness drug purchases, extensive margin



(c) Any drug purchases, intensive margin



(d) Sickness absences days, over 10 days

Notes: Coefficient for treatment group and 95% confidence intervals. Standard errors are clustered at individual level.

Figure 5.A4 Descriptive evidence of average retirement age for two cohorts.

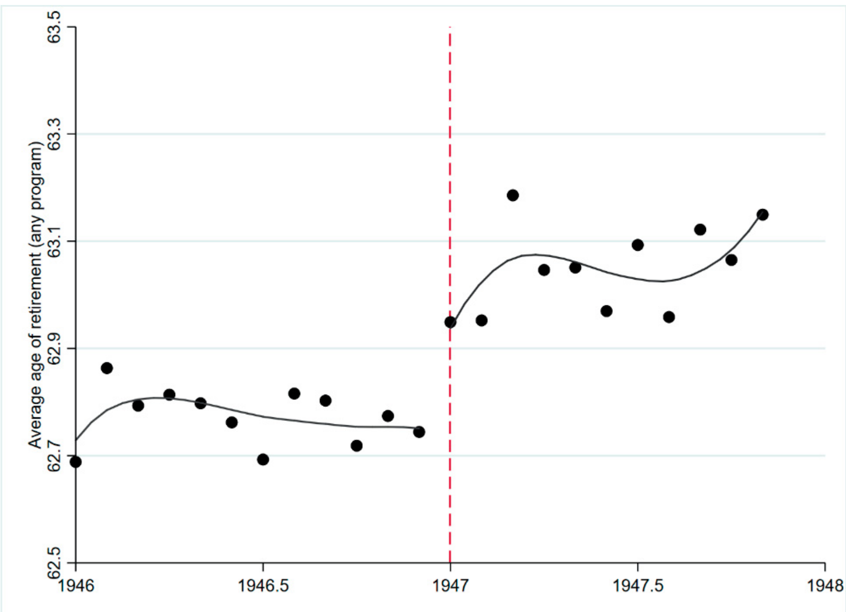
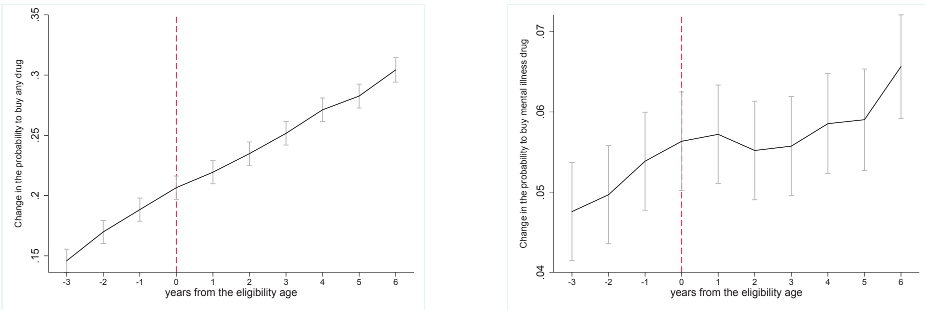
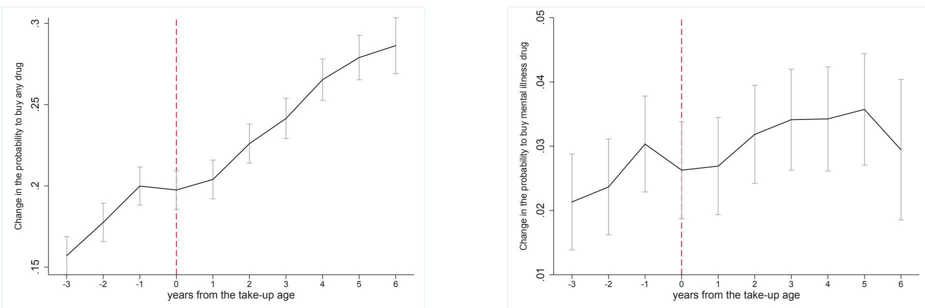


Figure 5.A5 Change in the probability to buy any medicine (left figure) or mental illness drug (right figure) during a year by distance from the eligibility age. Placebo, non-part-time pensioners.



Notes: the estimates are based on a fixed effects model where years from the eligibility act as explanatory variables. The estimation is done for years 1995-2004 for the total part-time pensioners sample. The vertical lines represents 95% confidence intervals.

Figure 5.A6 Change in the probability to buy any medicine (left figure) or mental illness drug (right figure) during a year by distance from the take-up year.



Notes: the estimates are based on a fixed effects model where years from the take-up act as explanatory variables. The estimation is done for years 1995-2004 for the total part-time pensioners sample. The vertical lines represents 95% confidence intervals.